

The Logarithm Map, its Limits and Fréchet Means in Orthant Spaces

D. Barden^{*} H. Le[†]

Abstract

The first part of the paper studies the expression for, and the properties of, the logarithm map on an orthant space, which is a simple stratified space, with the aim of analysing Fréchet means of probability measures on such a space. In the second part, we use these results to characterise Fréchet means and to derive various of their properties, including the limiting distribution of sample Fréchet means.

Keywords: Directional derivative; Fréchet mean; logarithm map; stratified space.

AMS MSC 2010: 60B05; 60B10.

1 Introduction

Several papers have recently appeared concerning probabilistic and statistical analysis of data on certain stratified spaces (cf. [5], [2], [10], [1] and [11]). One such example is the analysis of phylogenetic trees on the BHV space introduced in [5] (cf. [9], [19], [17], [3], [12], [15] and [18]). The BHV space \mathbf{T}_{m+2} of metric trees with $m + 2$ leaves is a stratified CAT(0)-space with each stratum being isometric with a positive Euclidean orthant that is at most m -dimensional. It is already clear from these preliminary results that some fundamental statistics exhibit strikingly different features from the corresponding ones on Euclidean spaces or on manifolds and that one faces significant challenges in developing novel tools to analyse them, on account of the non-trivial topological structure of these spaces. It also becomes apparent that, although the topological and geometrical properties of stratified spaces have been extensively studied and are mostly well understood, many of the properties required for probabilistic and statistical analysis of data on these spaces have not been addressed.

This paper concentrates on orthant spaces introduced in [15], a relatively simple type of stratified space but more general than the space \mathbf{T}_{m+2} of phylogenetic trees. The latter has $(2m + 1)!!$ m -dimensional strata, together with their bounding strata, selected from among the $\binom{M}{m}$ positive orthants in \mathbb{R}^M

^{*}Girton College, University of Cambridge, Cambridge, CB3 0JG, UK (d.barden@dpms.cam.ac.uk).

[†]School of Mathematical Sciences, University of Nottingham, Nottingham, NG7 2RD, UK (huiling.le@nottingham.ac.uk).

where $M = 2^{m+2} - m - 4$. In particular, each co-dimension one stratum bounds exactly three top-dimensional strata. Thus not only are the relevant dimensions sparse, but the percentage of the positive orthants occupied by the tree space of each dimension declines exponentially. These constraints, such as the restrictions on the dimension and the number of orthants involved in the space, no longer hold in a general orthant space, although we do have to make one restriction to ensure that it is a CAT(0)-space. We shall recall, in the next section, the concept of an orthant space, introducing the subsidiary concepts and definitions we use to describe the structure of such spaces and, in particular, of their tangent cones at the various points.

A fundamental concept for statistical analysis of non-Euclidean data is that of the Fréchet mean, which generalises the concept of the mean of Euclidean data. A point \mathbf{x}_0 in a metric space \mathbf{M} is a Fréchet mean of a probability measure μ on \mathbf{M} if, at \mathbf{x}_0 , the Fréchet function of μ defined by

$$\frac{1}{2} \int_{\mathbf{M}} d(\mathbf{x}, \mathbf{x}')^2 d\mu(\mathbf{x}') \quad (1)$$

attains its global minimum. In order to characterise and locate Fréchet means, we need to take directional derivatives of the Fréchet function and hence, implicitly, of the distance function. The latter involves the logarithm map $\log_{\mathbf{x}^*}(\mathbf{x})$ which, analogous to the inverse of the exponential map on manifolds, is the initial tangent vector to the geodesic from \mathbf{x}^* to \mathbf{x} . This logarithm map is globally well-defined on CAT(0)-spaces and has been studied, for example, in [14] and [16]. However, these results do not cover all the properties required for our analysis, although naturally we do rely on some of their results. On the other hand, an algorithm for finding the geodesic between any two given trees in the tree space \mathbf{T}_{m+2} was given in [19] and, using the analysis behind that algorithm, the expression for the logarithm map $\log_{\mathbf{x}^*}$ was obtained in [3] when \mathbf{x}^* lies in a top-dimensional stratum. Although, as pointed out in [15], this expression for $\log_{\mathbf{x}^*}$ could be extended to more general orthant spaces, it is noted in [3] that these results are not adequate to provide a tool for analysing Fréchet means when they lie in any stratum of co-dimension at least two. The latter requires a better understanding of the behaviour of the logarithm map as the end points of the geodesics move within and between strata. To this end, we first re-examine geodesics directly from first principles in Section 3, in particular avoiding the implicit assumption that \mathbf{x}^* lies in a top-dimensional stratum. This leads, in Theorem 1, to an explicit expression for a version of the logarithm map that we shall use, valid for any point in an orthant space. Since this expression is determined by the carrier of the geodesic, we analyse possible changes in that carrier, focussing on the set, specified in Definition 8, of points \mathbf{x} at which significant changes occur. This allows us, in Section 4, to derive the limit of the log map as the reference point \mathbf{x}' approaches \mathbf{x}^* from a co-bounding stratum. We also study the projections of these limits, and the limits of the projections, onto the various strata related to the stratum in which \mathbf{x}^* lies. This enables us to prove the existence of, and to identify, certain of their derivatives and directional derivatives.

With this understanding of the logarithm map, the second part of the paper turns its attention to the analysis of Fréchet means. In Section 5 we obtain, in Theorem 3, the necessary and sufficient conditions for a point \mathbf{x}^* to be the Fréchet mean of a probability measure on the orthant space \mathbf{X}^m . Two special

sets arise in this analysis. Firstly, one of the criteria in Theorem 3 involves an inequality and the set, specified in Definition 10, of vectors in the tangent cone to \mathbf{X}^m at the Fréchet mean for which that is an equality is significant. Secondly, there is the set given by Definition 11. This is related to a limit of the logarithm map and, in a certain sense, encapsulates the ‘departure’ of this limit from the analogous behaviour of the logarithm map on a Euclidean space. Both of these sets are related to the limiting distribution of sample Fréchet means, which we establish in the final Section 6. There, in particular, we relate the limiting distribution with Euclidean Gaussian random variables. The covariance matrices of these random variables are related to the derivative of the projection of the logarithm map and to projections of the limits of the logarithm map.

Although we do not make it explicit, in view of our previous results for \mathbf{T}_{m+2} and the comments in [15], our interest in this paper is primarily in the case that \mathbf{x}^* lies in a stratum of co-dimension at least two. The results, when restricted to a top-dimensional or co-dimension one stratum, do generalise those for tree spaces in [3] although the approach here is necessarily more complex in order to encompass all cases.

2 Orthant spaces

An orthant space is a union of orthants in a common Euclidean space with certain natural constraints that ensure, for example, that such spaces are also CAT(0). Orthant spaces were first introduced in [15] as a generalisation of the tree spaces of [5].

Definition 1. *For two given integers $M \geq m$, an orthant space \mathbf{X}^m of dimension m is a subspace of the Euclidean space \mathbb{R}^M which is a union of open positive orthants of maximum dimension m such that*

- (i) *for every orthant σ in \mathbf{X}^m , the orthants in the closure $\overline{\sigma}$ of σ are also included in \mathbf{X}^m ;*
- (ii) *if, for any positive orthant σ in \mathbb{R}^M , all the 2-dimensional orthants in its closure are in \mathbf{X}^m , then σ itself is in \mathbf{X}^m .*

Note that the first condition correlates with the constraints used in the definition for orthant space in [15] and the second one restricts attention to the ‘non-positively curved’ orthant spaces in [15] (Proposition 6.10). These two conditions were first used by the authors of [5] to ensure the CAT(0)-property for tree spaces. Note also that, since \mathbf{X}^m is a union of orthants in a fixed Euclidean space \mathbb{R}^M , the number of strata in \mathbf{X}^m is always finite.

The orthant space \mathbf{X}^m so defined is a Whitney stratified set in the sense of Thom, [20], the strata being the various orthants that comprise \mathbf{X}^m . It has the structure of a cone with vertex, or ‘cone point’, the origin o , since each orthant is such a cone without its vertex, but that vertex, the origin, is necessarily included in \mathbf{X}^m . Note however that, unlike [6], it will be more convenient for us to take our strata to be relatively open as specified in Definition 1.

The CAT(0)-property of the orthant space \mathbf{X}^m results from the following indicated results in [6]. Firstly, there is the *intrinsic* subset metric d on \mathbf{X}^m ,

referred to as the length metric in [6], for which the distance $d(\mathbf{x}_1, \mathbf{x}_2)$ is the infimum of the lengths of rectifiable curves joining \mathbf{x}_1 to \mathbf{x}_2 . On the other hand, the intersection L of \mathbf{X}^m with the unit sphere in \mathbb{R}^M , is a simplicial complex on account of condition (i) and, since the axes in \mathbb{R}^M are orthogonal, it is an ‘all-right spherical complex’ (Section 7A.10) which, on account of condition (ii), is a ‘flag complex’. Then, by a theorem of Gromov (Theorem 5.18), L is a CAT(1)-space. The metric on \mathbf{X}^m implied by describing it as the 0-cone over L (Definition 5.6) is the intrinsic metric so that, by the theorem of Berestowski (Theorem 3.14), \mathbf{X}^m is CAT(0).

In particular, by the Cartan-Hadamard theorem (cf. [6], p.193), there is a unique geodesic between any two points of the orthant space \mathbf{X}^m . It follows that each stratum is totally geodesic in the strong sense that, if a geodesic contains two points of a stratum, it must include the entire linear segment in that stratum determined by those two points. On the other hand, although the distance metric for the CAT(0)-structure is the induced Euclidean metric, the angles along and between curves may differ for the two approaches. For example, a geodesic, defined as a shortest path between its endpoints in either context, will be a piecewise linear curve in \mathbb{R}^M , linear in each stratum, with angle $\pi/2$ in the Euclidean subspace metric where it changes stratum. However, for the CAT(0)-structure, that angle is defined to be π .

Definition 2. *Given the standard ordered orthonormal basis $U = (u_1, \dots, u_M)$ of \mathbb{R}^M , for any subset E of U , the Euclidean subspace spanned by the vectors in E is denoted by $\mathbb{R}(E)$ and its (strictly) positive orthant by $\mathcal{O}(E)$.*

For $E \subseteq F \subseteq U$, $\mathcal{O}(E)$ is said to bound $\mathcal{O}(F)$ and $\mathcal{O}(F)$ to co-bound $\mathcal{O}(E)$.

Note that there is no loss of generality in restricting \mathbf{X}^m to contain only positive orthants: given two orthants that differ only in having positive or negative coordinates with respect to one particular axis, the intrinsic metric will be the same as it would be if we replace, say the negative axis, by an axis orthogonal to \mathbb{R}^M . Thus, rather than considering \mathbf{X}^m to be a union of arbitrary orthants in \mathbb{R}^M , we could consider it to be a union of *positive* orthants in \mathbb{R}^{2M} .

We should also note that, unlike the case for tree spaces, strata of lower dimension than m need not bound any higher dimensional strata. When they do not, we shall refer to such strata as being locally top-dimensional.

Throughout the rest of the paper, \mathbf{X}^m will denote an orthant space of fixed dimension m viewed as comprising strata that are positive orthants of a fixed Euclidean space \mathbb{R}^M , where M is not necessarily $2^{m+2} - m - 4$ as it would be for tree space. Also, whenever we specify an orthant by a union of subsets of U , that will always be intended as a union of mutually disjoint subsets.

The tangent cone

It is natural for our purposes to follow [6] and to define the tangent cone to \mathbf{X}^m at a point \mathbf{x} to consist of all initial tangent vectors to smooth curves starting from \mathbf{x} , the smoothness possibly only being one-sided at \mathbf{x} . Note, however, that this is not the same as the generalised tangent space of [8]. To describe the tangent cone in more detail we work in \mathbb{R}^M . Then, when \mathbf{x} lies in a top-dimensional, or locally top-dimensional, stratum σ of dimension $m' (\leq m)$, the orthant space \mathbf{X}^m is locally an m' -dimensional manifold so that a smooth curve

can be extended on both sides of \mathbf{x} . Thus, the tangent cone will be the usual tangent space, a subspace of \mathbb{R}^M isometric with $\mathbb{R}^{m'}$ and tangent to σ . However, if \mathbf{x} lies in a stratum of positive co-dimension that is not locally top-dimensional, the orthant space \mathbf{X}^m is no longer locally a manifold. Consequently, the tangent cone at \mathbf{x} is no longer a Euclidean space. For example, if the stratum σ has co-dimension one and bounds top-dimensional strata, the tangent cone to \mathbf{X}^m at \mathbf{x} is an open book with a page corresponding to each of the top-dimensional strata co-bounding σ and with the co-dimension one stratum σ being extended to \mathbb{R}^{m-1} to form its spine.

More generally, the tangent cone at a point \mathbf{x} , in a stratum $\sigma = \mathcal{O}(E)$ of co-dimension $l (\geq 1)$, has a stratification imitating that of \mathbf{X}^m itself in the neighbourhood of \mathbf{x} : for each stratum $\tau = \mathcal{O}(E \cup F)$ of co-dimension $l' < l$ that co-bounds σ , so that F comprises the vectors that have positive coordinates in τ but zero coordinates in σ , there is the stratum $\mathbb{R}(E) \times \mathcal{O}(F)$ in the tangent cone. Then, the tangent cone at \mathbf{x} has a stratification determined by identifying the various copies of $\mathbb{R}(E)$ as well as the tangent semi-axes shared by pairs of strata that co-bound σ . In particular, when no strata co-bound σ , the tangent cone is simply the Euclidean space $\mathbb{R}(E)$.

Definition 3. Let $\sigma = \mathcal{O}(E)$ and $\tau = \mathcal{O}(E \cup F)$ be two strata in \mathbf{X}^m with co-dimensions l and $l' < l$, respectively. The component $\mathbb{R}(E)$ common to all the strata in the tangent cone to \mathbf{X}^m at $\mathbf{x} \in \sigma$ is referred to as the tangent space to σ at \mathbf{x} . Vectors in the stratum $\mathbb{R}(E) \times \mathcal{O}(F)$ of the tangent cone at $\mathbf{x} \in \sigma$ with non-zero second component are referred to as vectors tangent to τ at \mathbf{x} .

The set of unit vectors in $\mathbb{R}(E) \times \mathcal{O}(F)$ is denoted by $\mathcal{S}_{\tau, \sigma}^{m-l'}$ and the set of those in $\{\mathbf{0}\} \times \mathcal{O}(F)$ by $\mathcal{S}_{\tau \setminus \sigma}^{l-l'}$.

Note that the basis vectors in E do not generally precede those of F in U , and so writing the stratum as $\mathbb{R}(E) \times \mathcal{O}(F)$ implies an appropriate permutation of the coordinates. In the following, we shall continue to take this permutation for granted and usually not refer to it explicitly.

Inherited from the CAT(0)-structure of \mathbf{X}^m , the tangent cone to \mathbf{X}^m at \mathbf{x} , since it is metrically complete, also has a CAT(0)-structure (cf. [6], Theorem 3.19). Although the CAT(0)-metric on \mathbf{X}^m is, by definition, the intrinsic subset metric, the CAT(0)-metric on the tangent cone to \mathbf{X}^m at \mathbf{x} is defined (cf. [16], p144) by

$$\rho_{\mathbf{x}}(\mathbf{w}_1, \mathbf{w}_2) = \{\|\mathbf{w}_1\|^2 + \|\mathbf{w}_2\|^2 - 2 \ll \mathbf{w}_1, \mathbf{w}_2 \gg\}^{1/2}$$

for tangent vectors \mathbf{w}_1 and \mathbf{w}_2 at \mathbf{x} with

$$\ll \mathbf{w}_1, \mathbf{w}_2 \gg = \|\mathbf{w}_1\| \|\mathbf{w}_2\| \cos \angle_{\mathbf{x}}(\mathbf{x}_1, \mathbf{x}_2), \quad (2)$$

where the points \mathbf{x}_1 and \mathbf{x}_2 lie on the geodesics from \mathbf{x} in directions \mathbf{w}_1 and \mathbf{w}_2 respectively and $\angle_{\mathbf{x}}(\mathbf{x}_1, \mathbf{x}_2)$ is the Alexandrov-angle at \mathbf{x} of the geodesic triangle $\Delta(\mathbf{x} \mathbf{x}_1 \mathbf{x}_2)$ (cf. [6], Section 1.12), which is independent of the choice of the points \mathbf{x}_1 and \mathbf{x}_2 . Note that, in general, \ll, \gg on the tangent cones should be distinguished from the usual Euclidean inner product \langle, \rangle . However, in the case of \mathbf{X}^m , any geodesic triangle contained in the closure of a stratum is in fact a Euclidean geodesic triangle and its angles are the Euclidean ones.

Hence, in particular, $\ll \mathbf{w}_1, \mathbf{w}_2 \gg = \langle \mathbf{w}_1, \mathbf{w}_2 \rangle$ for any $\mathbf{w}_1, \mathbf{w}_2$ in the closure of $\mathbb{R}(E) \times \mathcal{O}(F)$. In other words, the CAT(0)-metric on the tangent cone to \mathbf{X}^m at any point \mathbf{x} is again the same as the intrinsic one inherited from the tangent space to \mathbb{R}^M at \mathbf{x} .

3 The logarithm map

Analogous to an inverse of the exponential map on a Riemannian manifold, the logarithm map on \mathbf{X}^m is defined as follows.

Definition 4. *The logarithm map at $\mathbf{x}^* \in \mathbf{X}^m$ is the map $\log_{\mathbf{x}^*}(\mathbf{x})$ from \mathbf{X}^m to the tangent cone to \mathbf{X}^m at \mathbf{x}^* , the image of \mathbf{x} being the initial tangent vector, with norm $d(\mathbf{x}^*, \mathbf{x})$, to the geodesic from \mathbf{x}^* to \mathbf{x} .*

The logarithm map is globally well-defined since, as already mentioned, the Cartan-Hadamard theorem implies that there is a unique geodesic between any two points \mathbf{x}^* and \mathbf{x} of \mathbf{X}^m . If that geodesic has an initial segment in a stratum containing \mathbf{x}^* it will certainly have an initial tangent vector. If it has only \mathbf{x}^* in the initial stratum, it must then have an open segment $\gamma(0, \epsilon)$, with $\gamma(0) = \mathbf{x}^*$, in a co-bounding stratum. Then it will still have a one-sided derivative at \mathbf{x}^* which suffices to define the logarithm map.

In order to analyse the logarithm map, we first need to understand the geodesics. The intersection of a geodesic with a stratum, a Euclidean orthant, will be either a single point or a complete intersection of a Euclidean line with that orthant.

Definition 5. *The carrier of a geodesic is the sequence of strata each of whose intersection with the geodesic is a Euclidean line of positive length.*

This terminology was introduced in [21] in the context of tree spaces. The case of a single point intersection arises between successive strata of the carrier: between the (open) linear segment in one stratum and that in the next, there will be one point in the common bounding stratum of those two strata. This intermediate stratum is not listed in the carrier; it is in fact specified by the adjacent strata as the stratum of highest dimension in the intersection of their closures. Similarly, when a geodesic starts, or ends, in a stratum of positive co-dimension and does not remain in that stratum, but passes immediately to a co-bounding stratum, then the latter will be the first, or last, stratum in the carrier. In such a situation, we shall regard the point in the bounding stratum as having the same set of semi-axes as the co-bounding stratum, albeit with the relevant coordinates zero. That is, we regard it as a point of the closure of the co-bounding stratum.

To describe the carrier of the geodesic from \mathbf{x}^* to \mathbf{x} in more detail, as well as for later analysis, we require the following terminology, where we shall abuse the usual terminology and refer to the basis vectors in a set $E \subseteq U$ as the axes both of $\mathcal{O}(E)$ and of any point in $\mathcal{O}(E)$.

Definition 6. (i) *For any point \mathbf{x} in \mathbf{X}^m , $E_{\mathbf{x}}$ denotes the set of axes for which \mathbf{x} has positive coordinates and, for any axis $u \in U$, $|u|_{\mathbf{x}}$ denotes its u -coordinate.*

(ii) For a subset $A \subseteq U$, the number of axes in A is denoted by $|A|$; the set of axes in A is denoted by $(e_1, \dots, e_{|A|})$ ordered as in U ; the vector $(|e_1|_{\mathbf{x}}, \dots, |e_{|A|}|_{\mathbf{x}})$ is denoted by $A_{\mathbf{x}}$ and its norm by $\|A\|_{\mathbf{x}}$.

(iii) The subsets A and B of U are said to be compatible in the orthant space \mathbf{X}^m if the orthant $\mathcal{O}(A \cup B)$ is contained in \mathbf{X}^m . A and B are called totally incompatible in \mathbf{X}^m if no axis in A is compatible in \mathbf{X}^m with any axis in B .

To identify the carrier in terms of the axes involved we observe that, for a geodesic $\gamma(t)$, each coordinate function $(\gamma(t))_e$ must be linearly interpolated between any two values that are non-zero. It follows that once a particular coordinate, having been positive along the geodesic, becomes zero it must remain so or, having started at zero, once it becomes positive, it must continue monotonically to its final value. In particular, the only basis vectors that can occur with positive coordinate at any point along the geodesic from \mathbf{x}^* to \mathbf{x} are those that belong to \mathbf{x}^* or \mathbf{x} or to both. Moreover, any axis e appended, as above, to $E_{\mathbf{x}^*}$ by virtue of the geodesic from \mathbf{x}^* passing immediately to a co-bounding stratum must have coordinate zero at \mathbf{x}^* increasing linearly along the geodesic to its value at \mathbf{x} . Thus, e is in $E_{\mathbf{x}}$ and is compatible with all the axes in $E_{\mathbf{x}^*}$ and any such e must occur in this way. Then, the set of axes *common* to all strata along the geodesic comprises those that have positive coordinates in both \mathbf{x}^* and \mathbf{x} , together with axes $e^* \in E_{\mathbf{x}^*}$ compatible with $E_{\mathbf{x}}$ and axes $e \in E_{\mathbf{x}}$ compatible with $E_{\mathbf{x}^*}$. Note that this set is determined by $E_{\mathbf{x}^*}$ and $E_{\mathbf{x}}$ and so independent of where \mathbf{x}^* and \mathbf{x} lie in their (open) strata $\mathcal{O}(E_{\mathbf{x}^*})$ and $\mathcal{O}(E_{\mathbf{x}})$.

The number $k + 1$ of orthants in the carrier $\mathcal{C} = (\mathcal{O}_0, \mathcal{O}_1, \dots, \mathcal{O}_k)$ of the geodesic from \mathbf{x}^* to \mathbf{x} will, naturally, depend on both \mathbf{x} and \mathbf{x}^* . If \mathbf{x}^* lies in a top dimensional stratum it will have m strictly positive coordinates, all of which, assuming that none are also positive in \mathbf{x} , must become zero somewhere along the geodesic and at least one must become zero on each change of stratum as they cannot vanish within a stratum of the carrier. Thus, there will be $m + 1$ strata in the carrier, that is $k = m$, if and only if they vanish one at a time. So, $k < m$ if and only if somewhere along the geodesic at least two coordinates become zero on passing from \mathcal{O}_i to \mathcal{O}_{i+1} . When \mathbf{x}^* and \mathbf{x} have k_0 axes in common, the corresponding coordinates will remain positive throughout the intervening geodesic and the corresponding axes will belong to all the strata of the carrier, so that the maximum value of k would now be $k' = m - k_0$. Similarly, if \mathbf{x}^* were in a stratum of dimension m_0 , this maximum would be $m_0 - k_0$.

We shall denote the set of common axes of \mathbf{x}^* and \mathbf{x} by both A_0 and B_0 to accord with the following. Each member of the sequence of strata $\mathcal{C} = (\mathcal{O}_0, \mathcal{O}_1, \dots, \mathcal{O}_k)$ that comprise the carrier of the geodesic from \mathbf{x}^* to \mathbf{x} has $\mathcal{O}(A_0)$ as a factor. We write $\mathcal{C}' = (\mathcal{O}'_0, \mathcal{O}'_1, \dots, \mathcal{O}'_k)$ for the sequence with this common factor removed. It is the carrier for the projection of the original geodesic onto the orthogonal complement of $\mathbb{R}(A_0)$, now between two points with no common axes. This would be the original points, and the (re-parametrised) geodesic between them, if they had no common axis. Given the carrier of such a geodesic between points without common axes, it determines two sequences (A_1, \dots, A_k) and (B_1, \dots, B_k) of subsets of axes, where A_i is the set of all the axes whose coordinates become zero and B_i the set of all those whose coordinates become positive as the geodesic passes from \mathcal{O}'_{i-1} to \mathcal{O}'_i . Thus, the stratum \mathcal{O}'_i is the stratum $\mathcal{O}'_i = \mathcal{O}(B_1 \cup \dots \cup B_i \cup A_{i+1} \cup \dots \cup A_k)$,

with \mathcal{O}'_0 determined by $A_1 \cup \dots \cup A_k$. In particular, the sets B_i and A_j of axes are non-empty for all i and j and compatible in \mathbf{X}^m for $i < j$.

Definition 7. *The support of the geodesic from \mathbf{x}^* to \mathbf{x} is defined to be the pair $(\mathcal{A}, \mathcal{B})$ of sequences*

$$\mathcal{A} = (A_0, A_1, \dots, A_k) \quad \text{and} \quad \mathcal{B} = (B_0, B_1, \dots, B_k).$$

Then, the set $A_0 \cup A_1 \cup \dots \cup A_k$ is the full set of axes of the initial point \mathbf{x}^* of the geodesic, allowing for certain of these to have zero length if \mathbf{x}^* is in a stratum of positive co-dimension that the geodesic leaves immediately. Similarly, with a similar proviso, $B_0 \cup B_1 \cup \dots \cup B_k$ is the set of axes of its final point \mathbf{x} .

Since all axes common to \mathbf{x}^* and \mathbf{x} are contained in $A_0 = B_0$, $A_1 \cup \dots \cup A_k$ is disjoint from $B_1 \cup \dots \cup B_k$. This implies that $A_i \cap B_j = \emptyset$ for all i and j that are not both zero. *A fortiori* an axis once removed cannot be removed again, or once introduced cannot be introduced again, so that $A_i \cap A_j = \emptyset$ and $B_i \cap B_j = \emptyset$ for all $i \neq j$. In particular, the intermediate stratum between \mathcal{O}'_{i-1} and \mathcal{O}'_i , their common boundary component, is

$$\mathcal{O}(B_1 \cup \dots \cup B_{i-1} \cup A_{i+1} \cup \dots \cup A_k).$$

Although B_i is compatible with $A_{i+1} \cup \dots \cup A_k$ we see, in the following, that B_i is totally incompatible with A_i for $i > 0$.

Proposition 1. *For $i > 0$, no axis $a \in A_i$ is compatible with any axis $b \in B_i$, where A_i and B_i appear in the support $(\mathcal{A}, \mathcal{B})$ of the geodesic γ from \mathbf{x}^* to \mathbf{x} .*

Proof. The projection $\gamma_{a,b}$ of γ onto $\mathbb{R}(a, b)$, given by the a - and b -components of the points along γ , must go from the point $(|a|_{\mathbf{x}^*}, 0)$ to $(0, |b|_{\mathbf{x}})$. Since, by the definition of A_i and B_i , the A_i -coordinates along γ must vanish before the B_i -coordinates may grow, this projection has to go along the a -axis in $\mathbb{R}(a, b)$ to the origin and then along the b -axis. However, if a and b were compatible, the stratum $\mathcal{O}(a, b)$ would lie in \mathbf{X}^m and provide a shorter path $\tilde{\gamma}_{a,b}$. Since all the axes in \mathbf{X}^m are mutually orthogonal, replacing the component $\gamma_{a,b}$ of γ by $\tilde{\gamma}_{a,b}$ would produce a path $\tilde{\gamma}$ from \mathbf{x}^* to \mathbf{x} shorter than γ . \square

With the above description of the carrier of the geodesic γ from \mathbf{x}^* to \mathbf{x} , we may derive its initial tangent vector, or equivalently $\log_{\mathbf{x}^*}(\mathbf{x})$. As in [2] and [3] for the space of trees, it will be convenient to have a modified version of the logarithm map. For this, since the tangent cones at various points in σ are all parallel, we may parallel translate them to the cone point o , the origin in \mathbb{R}^M , to produce a common isometric copy \mathbf{C}_σ . Note that, although the origin corresponds to the cone point o of the orthant space \mathbf{X}^m , \mathbf{C}_σ is not the tangent cone at o of \mathbf{X}^m , neither being contained in the other, unless $\sigma = \{o\}$. Then, since the coordinate vector of the point \mathbf{x}^* , which we also denote by \mathbf{x}^* , lies in the common factor $\mathbb{R}(E)$ of all the strata of \mathbf{C}_σ , it makes sense to add it to $\log_{\mathbf{x}^*}(\mathbf{x})$ and the result

$$\Phi(\mathbf{x}; \mathbf{x}^*) = \log_{\mathbf{x}^*}(\mathbf{x}) + \mathbf{x}^*$$

will still lie in \mathbf{C}_σ . We shall also refer to Φ as the ‘logarithm map’ leaving it to the context and notation to imply to which version we are referring.

Theorem 1. For any points \mathbf{x}^* and \mathbf{x} in \mathbf{X}^m , let the sequences $\mathcal{A} = (A_0, \dots, A_k)$ and $\mathcal{B} = (B_0, \dots, B_k)$ of sets of axes form the support of the geodesic from \mathbf{x}^* to \mathbf{x} and j the map induced by mapping the sequence A_0, A_1, \dots, A_k of axes onto their standard representation in \mathbb{R}^M . Then, the logarithm map $\Phi(\cdot; \mathbf{x}^*)$ at \mathbf{x}^* is given by

$$\Phi(\mathbf{x}; \mathbf{x}^*) = j \left((B_0)_{\mathbf{x}}, -\frac{\|B_1\|_{\mathbf{x}}}{\|A_1\|_{\mathbf{x}^*}}(A_1)_{\mathbf{x}^*}, \dots, -\frac{\|B_k\|_{\mathbf{x}}}{\|A_k\|_{\mathbf{x}^*}}(A_k)_{\mathbf{x}^*} \right). \quad (3)$$

In particular, $\Phi(\cdot; \lambda \mathbf{x}^*) = \Phi(\cdot; \mathbf{x}^*)$ for any constant $\lambda > 0$.

Proof. The orthogonal projection of γ onto $\mathcal{O}(A_0)$ will have the lengths of the axes in A_0 , those common to \mathbf{x}^* and \mathbf{x} , of the points along γ linearly interpolated between their values in \mathbf{x}^* and in \mathbf{x} . This will determine the component of the initial tangent vector to γ that is tangent to $\mathcal{O}(A_0)$, namely

$$v_0 = (B_0)_{\mathbf{x}} - (A_0)_{\mathbf{x}^*} \in \mathbb{R}(A_0). \quad (4)$$

For the remaining coordinates, since the sets A_i and B_j above are all mutually disjoint, it follows that, for each i , the subspace $\mathbb{R}(A_i \cup B_i)$ is orthogonal to all $\mathbb{R}(A_j)$ and $\mathbb{R}(B_j)$ for $j \neq i$, so that the coordinates of the geodesic γ that are in $\mathbb{R}(A_i \cup B_i)$ are just those of the projection γ_i of γ onto that subspace. Since A_i and B_i are totally incompatible, this projection must lie in $\mathcal{O}(A_i) \cup \mathcal{O}(B_i)$. Indeed, if s_i is the parameter such that $\gamma(s_i) \in \mathcal{O}_{i-1} \cap \mathcal{O}_i$, then $(A_i)_{\gamma(s)} \in \mathcal{O}(A_i)$ declines linearly from $(A_i)_{\gamma(0)} = (A_i)_{\mathbf{x}^*}$ to $(A_i)_{\gamma(s_0)} = \mathbf{0}$. Then, the coordinates $(B_i)_{\gamma(s)} \in \mathcal{O}(B_i)$ increases linearly from zero at $\gamma(s_i)$ to $(B_i)_{\gamma(1)} = (B_i)_{\mathbf{x}}$. Thus, the projected geodesic γ_i has length $\|A_i\|_{\mathbf{x}^*} + \|B_i\|_{\mathbf{x}}$ and its initial tangent vector is parallel to $-(A_i)_{\mathbf{x}}$, so that the initial tangent vector to γ_i is

$$v_i = -\frac{\|A_i\|_{\mathbf{x}^*} + \|B_i\|_{\mathbf{x}}}{\|A_i\|_{\mathbf{x}^*}}(A_i)_{\mathbf{x}^*}.$$

Hence, the initial tangent vector to γ with norm $d(\mathbf{x}^*, \mathbf{x})$ is represented by (v_0, v_1, \dots, v_k) . However, this ordering of the coordinates, with those in $\mathbb{R}(A_i)$ preceding those of $\mathbb{R}(A_{i+1})$ for each i , requires a shuffle of the coordinates before inserting the necessary zeros to obtain its representation with respect to the standard basis in \mathbb{R}^M . If we denote this linear process by j , then the logarithm map at \mathbf{x}^* will be

$$\log_{\mathbf{x}^*} : \mathbf{x} \mapsto j(v_0, v_1, \dots, v_k).$$

Then, (3) follows from the coordinates v_i since the coordinates of \mathbf{x}^* are

$$j((A_0)_{\mathbf{x}^*}, (A_1)_{\mathbf{x}^*}, \dots, (A_k)_{\mathbf{x}^*}).$$

□

Note that, when \mathbf{X}^m is a tree space and \mathbf{x}^* lies in a top-dimensional stratum, the expression for $\Phi(\mathbf{x}; \mathbf{x}^*)$ obtained above bears a similarity with the similarly modified logarithm map in [3] for the tree spaces, where $\Phi(\mathbf{x}; \mathbf{x}^*)$ was denoted by $\Phi_{\mathbf{x}^*}(\mathbf{x})$ and the permutation map π there corresponds to j here. This confirms the claim in Section 6 of [15] that the results on tree spaces obtained earlier in that paper also hold for orthant spaces.

The potential variation of the form that the algebraic expression (3) for the logarithm map takes, arising from the changes in the supports of the geodesics, is one of the main obstructions to generalising the theory for manifolds to or-thant spaces, or more general stratified spaces. However, certain changes in the supports do not affect the final expression (3) and we show first that, if the first components of the supports of two nearby geodesics differ only in a permutation of the subsets involved, then the second components of the supports will necessarily under go the same permutation.

Proposition 2. *For $\mathbf{x}_1^*, \mathbf{x}_2^*$ in the stratum τ_1 , and \mathbf{x}_1 and \mathbf{x}_2 in the stratum τ_2 , let $(\mathcal{A}^i, \mathcal{B}^i)$ denote the supports of the geodesics from \mathbf{x}_i^* to \mathbf{x}_i . If there is a permutation π of indices such that $\mathcal{A}^2 = \mathcal{A}_\pi^1$ (i.e. $A_i^2 = A_{\pi(i)}^1$), then $\mathcal{B}^2 = \mathcal{B}_\pi^1$.*

Proof. Since the components $A_0^i = B_0^i$ of the common axes in each of these two supports are determined by the relationship between the axes of τ_1 and τ_2 , they are invariant with respect to the positions of \mathbf{x}_i^* in τ_1 and \mathbf{x}_i in τ_2 , so that we must have $A_0^1 = A_0^2$ and any permutation π such that $\mathcal{A}^2 = \mathcal{A}_\pi^1$ must fix the index 0.

Assume that $\mathcal{A}^1 = (A_0^1, A_1^1, \dots, A_k^1)$, that $k > 1$ and that there is some $0 \leq m < k - 1$ such that

$$B_i^2 = B_{\pi(i)}^1 \quad \text{for } i \leq m.$$

We show that $B_{m+1}^2 = B_{\pi(m+1)}^1$ by contradiction.

If $B_{m+1}^2 \setminus B_{\pi(m+1)}^1 \neq \emptyset$, then there exists $1 \leq l \leq k$ such that $B_{m+1}^2 \cap B_l^1 \neq \emptyset$. Since B_i^2 are mutually disjoint, we have $l \neq \pi(1), \dots, \pi(m)$ by the assumption. If $l < \pi(m+1)$ then, since B_l^1 is compatible with $A_{l+1}^1 \cup \dots \cup A_k^1 \supseteq A_{\pi(m+1)}^1$, $B_{m+1}^2 \cap B_l^1$ is compatible with $A_{\pi(m+1)}^1 = A_{m+1}^2$, contradicting Proposition 1. However, if $l > \pi(m+1)$, then

$$A_l^1 \subseteq A_{\pi(m+2)}^1 \cup \dots \cup A_{\pi(k)}^1 = A_{m+2}^2 \cup \dots \cup A_k^2.$$

Since B_{m+1}^2 is compatible with $A_{m+2}^2 \cup \dots \cup A_k^2$, $B_l^1 \cap B_{m+1}^2$ is compatible with A_l^1 , again contradicting Proposition 1. Hence, we must have $B_{m+1}^2 \setminus B_{\pi(m+1)}^1 = \emptyset$.

To show that $B_{m+1}^2 = B_{\pi(m+1)}^1$ we note that $\mathcal{A}^1 = \mathcal{A}_{\pi^{-1}}^2$. Then, exchanging the roles of \mathcal{A}_1 and \mathcal{A}_2 in the above argument, we get $B_{\pi(m+1)}^1 \setminus B_{m+1}^2 = \emptyset$.

The required result follows by induction. \square

This result shows that, under hypotheses stated, if the sets A_i are permuted on passing from one geodesic to its neighbour then the norm $\|B_i\|_{\mathbf{x}}$ of the corresponding vector in the expression (3) will be carried along with it. Since the permutation j then takes all coordinates into their standard position, the net result will be unchanged.

There are other changes in carrier that do affect the form of the algebraic expression (3). These occur in occasional discrete steps as \mathbf{x} varies with \mathbf{x}^* fixed.

Proposition 3. *Let $(\mathcal{A}, \mathcal{B})$ be the support of the geodesic from \mathbf{x}^* to \mathbf{x}_1 .*

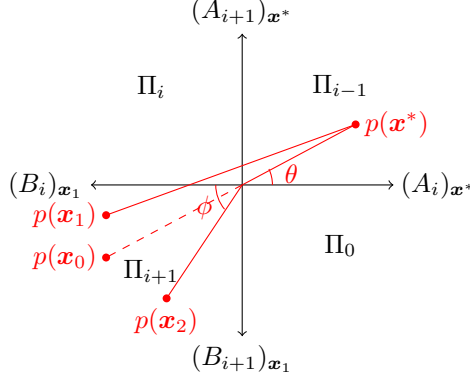


Figure 1: Change of carrier

(i) For all $i > 0$,

$$\frac{\|A_{i+1}\|_{x^*}}{\|A_i\|_{x^*}} > \frac{\|B_{i+1}\|_x}{\|B_i\|_x} \quad (5)$$

when $x = x_1$.

(ii) As x moves from x_1 within its stratum to a (first) point x_0 where, for some i , (5) becomes the equality

$$\frac{\|A_{i+1}\|_{x^*}}{\|A_i\|_{x^*}} = \frac{\|B_{i+1}\|_{x_0}}{\|B_i\|_{x_0}}, \quad (6)$$

the resulting algebraic expression (3) remains constant up to and including x_0 .

(iii) For $i > 0$, the sets A_i and B_{i+1} are either compatible or totally incompatible. In the former case, the algebraic expression (3) is unaffected as x moves from x_0 to points x_2 where the inequality (5) is reversed. However, in the latter case, that expression will change.

Proof. (i) We focus on three consecutive strata of the carrier, $\mathcal{O}_{i-1}, \mathcal{O}_i, \mathcal{O}_{i+1}$, projecting everything onto the subspace $\mathbb{R}(A_i \cup A_{i+1} \cup B_i \cup B_{i+1})$. As the geodesic passes from \mathcal{O}_{i-1} to \mathcal{O}_i , the coordinates along the axes in A_i become zero and those in B_i start to grow. Then, on passing from \mathcal{O}_i to \mathcal{O}_{i+1} , the coordinates of axes in A_{i+1} become zero and those in B_{i+1} grow. Consider the projection p of the geodesic onto the three planar quadrants Π_{i-1} determined by the vectors $(A_i)_{x^*}$ and $(A_{i+1})_{x^*}$, Π_i determined by $(A_{i+1})_{x^*}$ and $(B_i)_{x_1}$ and Π_{i+1} determined by $(B_i)_{x_1}$ and $(B_{i+1})_{x_1}$ as represented in Figure 1. This is an isometric representation of the relevant quadrants except that, in \mathbb{R}^M , all four vectors are mutually orthogonal. Then, \mathcal{O}_i is in the carrier if and only if the projection of the geodesic passes through Π_i . That is if and only if the angle θ that the vector $p(x^*)$ makes with the axis $(A_i)_{x^*}$ in Π_{i-1} is greater than the angle ϕ that $p(x_1)$ makes with $(B_i)_{x_1}$ in Π_{i+1} as expressed by (5) for $x = x_1$.

(ii) As the point \mathbf{x} moves from \mathbf{x}_1 to \mathbf{x}_2 , the length of the segment of the projection of the geodesic that lies in Π_i decreases to zero. This occurs at \mathbf{x}_0 when the angles θ and ϕ are equal. That is when (6) holds. Then, while the carrier of the geodesic from \mathbf{x}^* to \mathbf{x}_1 includes the sequence of strata \mathcal{O}_{i-1} , \mathcal{O}_i , \mathcal{O}_{i+1} , the geodesic from \mathbf{x}^* to \mathbf{x}_0 jumps immediately from \mathcal{O}_{i-1} to \mathcal{O}_{i+1} . The effect on the supports of the geodesics as their projections pass from \mathbf{x}_1 to \mathbf{x}_0 is that we replace the adjacent sets A_i , A_{i+1} and B_i , B_{i+1} in the sequences \mathcal{A} and \mathcal{B} that form the support of the former by $A_i \cup A_{i+1}$ and $B_i \cup B_{i+1}$ respectively in the support of the latter. Note however that the change of the carrier from \mathbf{x}_1 to \mathbf{x}_0 does not affect the expression for $\log_{\mathbf{x}^*}(\mathbf{x}_0)$ or $\Phi(\mathbf{x}_0; \mathbf{x}^*)$ since

$$\left(\frac{\|B_i\|_{\mathbf{x}_0}}{\|A_i\|_{\mathbf{x}^*}}(A_i)_{\mathbf{x}^*}, \frac{\|B_{i+1}\|_{\mathbf{x}_0}}{\|A_{i+1}\|_{\mathbf{x}^*}}(A_{i+1})_{\mathbf{x}^*} \right) = \frac{\|(B_i, B_{i+1})\|_{\mathbf{x}_0}}{\|(A_i, A_{i+1})\|_{\mathbf{x}^*}}(A_i, A_{i+1})_{\mathbf{x}^*}$$

in view of the equality (6) and the mutual orthogonality of all the axes.

(iii) As \mathbf{x} passes from \mathbf{x}_0 to \mathbf{x}_2 , there are two possibilities for the change of carrier. If the axis sets A_i and B_{i+1} are totally incompatible, so that neither the quadrant Π_0 determined by $(A_i)_{\mathbf{x}^*}$ and $(B_{i+1})_{\mathbf{x}}$, nor any stratum containing it lies in \mathbf{X}^m , then the carrier and its support will remain as they were for \mathbf{x}_0 and the algebraic expression (3) for the logarithm map will change as the equality (6), and hence the above equality, no longer holds for such \mathbf{x}_2 .

If, on the other hand, A_i and B_{i+1} are compatible sets of axes the projection of the geodesic from \mathbf{x}^* to \mathbf{x}_2 will be a straight line from $p(\mathbf{x}^*)$ to $p(\mathbf{x}_2)$ passing through the quadrant Π_0 . The effect on the support will be to transpose A_i and A_{i+1} in \mathcal{A} and B_i and B_{i+1} in \mathcal{B} : the coordinates of the axes in A_{i+1} vanish first and those in B_{i+1} grow first; \mathcal{O}_i becomes

$$\mathcal{O}(A_0 \cup B_1 \cup \dots \cup B_{i-1} \cup B_{i+1} \cup A_i \cup A_{i+2} \cup \dots \cup A_k).$$

However, as for the situation in Proposition 2, this reordering of the strata will be compensated for by a different j' instead of j , so that the resulting expressions for $\log_{\mathbf{x}^*}(\cdot)$ and $\Phi(\cdot; \mathbf{x}^*)$ remain locally constant, as should be expected since the four quadrants project isometrically onto a plane.

When \mathbf{x} passes from \mathbf{x}_2 to \mathbf{x}_1 in the reverse direction, the second possibility has a similar effect on the support transposing two consecutive subsets. However, in the first case, the reverse move would see A_i and B_{i+1} written as non-trivial disjoint unions $C_1 \cup C_2$ and $D_1 \cup D_2$ respectively where, for the relevant points, $\frac{\|C_1\|}{\|D_1\|} < \frac{\|C_2\|}{\|D_2\|}$. Note that, *a priori*, there is a further range of possibilities as \mathbf{x}_0 passes to \mathbf{x}_2 : if A_i and B_{i+1} are not totally incompatible, A_i could split as $C_1 \cup C_2$ and B_{i+1} as $D_1 \cup D_2$ with $C_1 \cup D_1$ contained in a stratum of \mathbf{X}^m and the geodesic from $p(\mathbf{x}^*)$ to $p(\mathbf{x}_2)$ passing through the stratum $\mathcal{O}(C_1 \cup D_1)$. However \mathbf{x}_0 would then have to satisfy both (6), that is

$$\|A_{i+1}\|_{\mathbf{x}^*} \|B_i\|_{\mathbf{x}_0} = \|C_1 \cup C_2\|_{\mathbf{x}^*} \|D_1 \cup D_2\|_{\mathbf{x}_0},$$

and its analogue for the alternative route

$$\|A_{i+1} \cup C_2\|_{\mathbf{x}^*} \|B_i \cup D_2\|_{\mathbf{x}_0} = \|C_1\|_{\mathbf{x}^*} \|D_1\|_{\mathbf{x}_0};$$

these equations are incompatible unless $C_2 = D_2 = \emptyset$. □

The carrier of the geodesic from \mathbf{x}^* to \mathbf{x} will also change when \mathbf{x} moves from one stratum to another which necessarily involves, as initial, final or intermediate stratum, a stratum of positive co-dimension. The set of all such strata, together with the quadratic hyper-surfaces determined by (6), form the defining boundaries for the *(pre)-vital polyhedral subdivision*, with respect to \mathbf{x}^* , in [15]. The points in any component of the complement of these surfaces all have the same carrier. However, since we shall only be concerned with changes in the forms of the algebraic expressions taken by $\log_{\mathbf{x}^*}(\mathbf{x})$, or equivalently by $\Phi(\mathbf{x}; \mathbf{x}^*)$, rather than changes in the underlying carrier, we shall focus on those changes of carrier that result from a stratum that is omitted not being replaced, not from two strata being replaced by two others. We encapsulate this in the following definition.

Definition 8. *Given a point $\mathbf{x}^* \in \mathbf{X}^m$, $\mathcal{D}_{\mathbf{x}^*}$ denotes the set that consists of all points $\mathbf{x} \in \mathbf{X}^m$ for which the support $(\mathcal{A}, \mathcal{B})$, where $\mathcal{A} = (A_0, \dots, A_k)$ and $\mathcal{B} = (B_0, \dots, B_k)$, of the geodesic from \mathbf{x}^* to \mathbf{x} has the property that, for one or more $i > 0$, the equality (6) holds with the sets of axes A_i and B_{i+1} totally incompatible.*

In view of the symmetry that reverses the geodesics at the same time as it reverses the order of the strata and interchanges the roles of the sequences \mathcal{A} and \mathcal{B} of edge sets in the support, the definition is symmetric: $\mathbf{x} \in \mathcal{D}_{\mathbf{x}^*}$ if and only if $\mathbf{x}^* \in \mathcal{D}_{\mathbf{x}}$. Since each stratum is a Euclidean orthant, it is preserved under multiplication by $\lambda > 0$ in \mathbb{R}^M which also multiplies the length of each curve by λ . Then, since the geodesic γ joining \mathbf{x}^* to \mathbf{x} is the shortest curve through the strata of \mathbf{X}^m from \mathbf{x}^* to \mathbf{x} , it follows that γ is mapped onto the geodesic from $\lambda \mathbf{x}^*$ to $\lambda \mathbf{x}$. In particular, these two geodesics have the same carrier. Thus, $\mathcal{D}_{\lambda \mathbf{x}^*} = \lambda \mathcal{D}_{\mathbf{x}^*}$ and, since the equations (6) are homogeneous, $\mathcal{D}_{\mathbf{x}^*} = \lambda \mathcal{D}_{\mathbf{x}^*}$.

For fixed \mathbf{x}^* , we may write \mathbf{X}^m as the finite union of sets \mathcal{X}_i , $i = 1, \dots, l$, where, for each i , \mathcal{X}_i comprises the points \mathbf{x} that have a given algebraic expression (3) for $\Phi(\mathbf{x}; \mathbf{x}^*)$. That is, each \mathcal{X}_i contains all points \mathbf{x} such that the sequence \mathcal{A} in the carriers of the geodesics between \mathbf{x}^* and \mathbf{x} is the same up to ordering of the subsets. Hence, for all pairs $i \neq j$, $\mathcal{X}_i \cap \mathcal{X}_j \subseteq \mathcal{D}_{\mathbf{x}^*}$. This pseudo-partition of \mathbf{X}^m with respect to \mathbf{x}^* determined by $\mathcal{D}_{\mathbf{x}^*}$ gives rise to a polyhedral subdivision of each stratum by restriction. It is coarser than the (pre)-vital subdivision of [15] and, if \mathbf{X}^m is a tree space and if \mathbf{x}^* lies in a top-dimensional stratum, it is equivalent to the polyhedral subdivision defined in [3].

4 Limits, projections and derivatives

We now turn to certain limits and projections of the logarithm map that, in particular, will enable us to calculate the directional derivatives we require.

Firstly, we obtain an expression for the limit of the logarithm map as the reference point \mathbf{x}^* moves along a geodesic. For a vector \mathbf{w} in the tangent cone at \mathbf{x}^* , write $\mathbf{x}^*(\lambda, \mathbf{w})$ for the point distant $\lambda \|\mathbf{w}\|$ along the geodesic γ starting at \mathbf{x}^* with initial tangent vector \mathbf{w} . Then, we have the following result.

Theorem 2. *Let $\sigma = \mathcal{O}(E)$ be a stratum of \mathbf{X}^m , $\mathbf{x}^* \in \sigma$ and \mathbf{x} be a fixed choice of point anywhere in \mathbf{X}^m .*

(i) If $\mathbf{w} \in \mathbb{R}(E)$ is tangent to σ at \mathbf{x}^* , then

$$\lim_{\lambda \rightarrow 0+} \Phi(\mathbf{x}; \mathbf{x}^*(\lambda, \mathbf{w})) = \Phi(\mathbf{x}; \mathbf{x}^*).$$

(ii) If σ bounds $\tau = \mathcal{O}(E \cup F)$ in \mathbf{X}^m and $\mathbf{w}_\tau \in \mathbb{R}(E) \times \mathcal{O}(F)$ is tangent to τ at \mathbf{x}^* , then the limit

$$\Psi(\mathbf{x}, \mathbf{w}_\tau; \mathbf{x}^*) = \lim_{\lambda \rightarrow 0+} \Phi(\mathbf{x}; \mathbf{x}^*(\lambda, \mathbf{w}_\tau))$$

exists. Moreover, there exist $\epsilon > 0$ and sequences $\mathcal{A} = (A_0, A_1, \dots, A_k)$ and $\mathcal{B} = (B_0, B_1, \dots, B_k)$ of sets of axes such that, for each $\lambda \in (0, \epsilon)$, $(\mathcal{A}, \mathcal{B})$ forms the support of the geodesic from $\mathbf{x}^*(\lambda, \mathbf{w}_\tau)$ to \mathbf{x} . In terms of these \mathcal{A} and \mathcal{B} ,

$$\Psi(\mathbf{x}, \mathbf{w}_\tau; \mathbf{x}^*) = j \left((B_0)_{\mathbf{x}}, -\frac{\|B_1\|_{\mathbf{x}}}{\|W_1\|} W_1, \dots, -\frac{\|B_k\|_{\mathbf{x}}}{\|W_k\|} W_k \right), \quad (7)$$

where $W_i = (A_i \cap E)_{\mathbf{x}^*}$, unless $(A_i \cap E)_{\mathbf{x}^*} = 0$, in which case $W_i = (A_i \cap F)_{\mathbf{w}_\tau}$, the projection of \mathbf{w}_τ on $\mathbb{R}(A_i)$, and j maps a vector to its standard representation in \mathbb{R}^M .

Proof. (i) This follows from the uniform continuity of geodesics with respect to their end points (cf. [6], pp195-196) and also from a minor modification of the proof of (ii) below.

(ii) Writing γ_λ for the geodesic from $\mathbf{x}^*(\lambda, \mathbf{w}_\tau)$ to \mathbf{x} , as $\mathbf{x}^*(\lambda, \mathbf{w}_\tau)$ moves along γ the support of γ_λ can only change when γ meets transversally one or more of the hyper-surfaces where the carrier of the geodesic to \mathbf{x} changes. This can only happen at discrete points along γ so, for some $\epsilon > 0$ and $0 < \lambda \leq \epsilon$, the carriers of the geodesics γ_λ will be independent of λ . Let $(\mathcal{A}, \mathcal{B})$ be the support of γ_ϵ from $\mathbf{x}^*(\epsilon, \mathbf{w}_\tau)$ to \mathbf{x} , where $\mathcal{A} = (A_0, A_1, \dots, A_k)$ and $\mathcal{B} = (B_0, B_1, \dots, B_k)$ and where $A_0 = B_0$ may be empty. Then, for $0 < \lambda \leq \epsilon$, the integer k and the support $(\mathcal{A}, \mathcal{B})$ will remain constant for the expression

$$\begin{aligned} & \Phi(\mathbf{x}; \mathbf{x}^*(\lambda, \mathbf{w}_\tau)) \\ &= j \left((B_0)_{\mathbf{x}}, -\frac{\|B_1\|_{\mathbf{x}}}{\|A_1\|_{\mathbf{x}^*(\lambda, \mathbf{w}_\tau)}} (A_1)_{\mathbf{x}^*(\lambda, \mathbf{w}_\tau)}, \dots, -\frac{\|B_k\|_{\mathbf{x}}}{\|A_k\|_{\mathbf{x}^*(\lambda, \mathbf{w}_\tau)}} (A_k)_{\mathbf{x}^*(\lambda, \mathbf{w}_\tau)} \right), \end{aligned}$$

replacing \mathbf{x}^* in (3) by $\mathbf{x}^*(\lambda, \mathbf{w}_\tau)$. This in particular implies that, for all $0 < \lambda \leq \epsilon$, the corresponding inequality (5) holds for all $i > 0$. Then, since the $\mathbf{x}^*(\lambda, \mathbf{w}_\tau)$ lie in τ for all sufficiently small positive λ , the vectors $\Phi(\mathbf{x}; \mathbf{x}^*(\lambda, \mathbf{w}_\tau))$ all lie in \mathcal{C}_τ so that it makes sense to take the limit as $\lambda \rightarrow 0+$.

To evaluate it, we take the limit in the above expression for $\Phi(\mathbf{x}; \mathbf{x}^*(\lambda, \mathbf{w}_\tau))$. Since $\mathbf{x}^* \in \mathcal{O}(E)$, $\mathbf{x}^*(\lambda, \mathbf{w}_\tau) = \mathbf{x}^* + \lambda \mathbf{w}_\tau$ for sufficiently small $\lambda > 0$ and it follows that $(A_i)_{\mathbf{x}^*(\lambda, \mathbf{w}_\tau)} = (A_i \cap E)_{\mathbf{x}^*} + \lambda (A_i)_{\mathbf{w}_\tau}$. So the limit as $\lambda \rightarrow 0+$ of this term is $(A_i \cap E)_{\mathbf{x}^*}$ if that is non-zero. If it is zero, then $A_i \cap E = \emptyset$ since $(e)_{\mathbf{x}^*} > 0$ for all $e \in E$. Then $(A_i)_{\mathbf{w}_\tau}$, the projection of \mathbf{w}_τ on $\mathbb{R}(A_i)$ is, in fact, $(A_i \cap F)_{\mathbf{w}_\tau}$. \square

If σ has co-dimension l and τ co-dimension l' then, when $l - l' = 1$ and so $|F| = 1$, there is no $i > 0$ such that $|A_i| > 1$ and $(A_i \cap E)_{\mathbf{x}^*} = 0$ as all the axes

involved in the carrier that are not in $E \cup F$ are in $A_0 = B_0$. If further $l = 1$ and $l' = 0$, that is, σ is a stratum of co-dimension one and τ co-bounding σ is a top-dimensional stratum, then $\Psi(\cdot, \mathbf{w}_\tau; \mathbf{x}^*)$ obtained here is identical with the map resulting from the ‘folding map’ composed with $\Phi(\cdot; \mathbf{x}^*)$ used in [3] when \mathbf{X}^m is a tree space, noting that \mathbf{w}_τ in this case is unique up to a positive scalar multiple.

For \mathbf{w}_τ as given in Theorem 2(ii), writing \mathbf{w}_τ^\perp for the component of \mathbf{w}_τ orthogonal to σ , that is, the component in $\{\mathbf{0}\} \times \mathcal{O}(F) \subset \mathbb{R}(E) \times \mathcal{O}(F)$, we observe the following consequences of Theorem 2, which imply that it suffices to restrict attention to $\mathbf{w}_\tau \in \mathcal{S}_{\tau \setminus \sigma}^{l-l'}$.

Corollary 1. *With the notation and hypotheses of Theorem 2(ii),*

(i) $\Psi(\cdot, \lambda \mathbf{w}_\tau; \mathbf{x}^*) = \Psi(\cdot, \mathbf{w}_\tau; \mathbf{x}^*)$ for all $\lambda > 0$;

(ii) $\Psi(\mathbf{x}, \mathbf{w}_\tau; \mathbf{x}^*) = \Psi(\mathbf{x}, \mathbf{w}_\tau^\perp; \mathbf{x}^*)$.

Proof. (i) is obvious and (ii) is immediate since $\sigma = \mathcal{O}(E)$, $\tau = \mathcal{O}(E \cup F)$ and only the F -coordinates of \mathbf{w}_τ are potentially involved in (7). \square

When \mathbf{x}^* lies in a stratum σ of positive co-dimension that is not locally top-dimensional, the vector $\log_{\mathbf{x}^*}(\mathbf{x})$ will usually have non-zero components both tangent to σ and orthogonal to it. In order to discuss the projections, and their derivatives, onto these components, we make the following definitions.

Definition 9. *For the stratum $\sigma = \mathcal{O}(E)$ in \mathbf{X}^m and any stratum $\tau = \mathcal{O}(E \cup F)$ co-bounding σ in \mathbf{X}^m , P_σ and $P_{\tau \setminus \sigma}$ respectively are the projections onto the two factors of the corresponding stratum $\mathbb{R}(E) \times \mathcal{O}(F)$ in \mathbf{C}_σ , or equivalently in the tangent cone at a point of σ , depending on the context.*

More generally, since each stratum in \mathbf{C}_τ has $\mathbb{R}(E \cup F) = \mathbb{R}(E) \times \mathbb{R}(F)$ as a factor, we shall also denote by P_τ , P_σ and $P_{\tau \setminus \sigma}$ the projections from \mathbf{C}_τ onto the factors $\mathbb{R}(E \cup F)$, $\mathbb{R}(E)$ and $\mathbb{R}(F)$ respectively.

For \mathbf{x}^* in σ or in τ co-bounding σ , we shall denote $P_\sigma(\log_{\mathbf{x}^*}(\mathbf{x}))$ by $\log_{\mathbf{x}^*}^\sigma(\mathbf{x})$ and $P_\sigma(\Phi(\mathbf{x}; \mathbf{x}^*))$ by $\Phi_\sigma(\mathbf{x}; \mathbf{x}^*)$. Note that, on \mathbf{C}_σ , P_σ so defined is the tangential projection onto σ and $P_{\tau \setminus \sigma}$ is one of several possible normal projections. In particular, for \mathbf{w}_τ above, $\mathbf{w}_\tau^\perp = P_{\tau \setminus \sigma}(\mathbf{w}_\tau)$. In the following, we shall extend the notation P_σ to include top-dimensional, or locally top-dimensional, strata by taking it to be the identity in that case, so that in particular $\Phi_\sigma(\mathbf{x}; \mathbf{x}^*) = \Phi(\mathbf{x}; \mathbf{x}^*)$ if σ is a top-dimensional, or locally top-dimensional, stratum.

For \mathbf{x}^* in σ of positive co-dimension, the non-zero components of $\log_{\mathbf{x}^*}(\mathbf{x})$ orthogonal to σ correspond to axes that have zero coefficient in \mathbf{x}^* and hence have non-zero coefficient in \mathbf{x} . Since these components are components of the initial tangent vector to the geodesic from \mathbf{x}^* to \mathbf{x} , the coefficient with respect to these axes must become non-zero and remain so throughout the geodesic. Hence, these axes are in $A_0 = B_0$, that is, they correspond to components of v_0 in (4). This implies, in particular, that $\Phi_\sigma(\mathbf{x}; \mathbf{x}^*)$ is given by (3) with $(B_0)_x$ there replaced by $(B_0 \cap E)_x$. Then, since the restriction to each stratum of the set \mathcal{D}_x defined by Definition 8 is relatively closed, the algebraic expression for $\Phi(\mathbf{x}; \mathbf{x}')$ will remain constant for \mathbf{x}' varying in a neighbourhood of \mathbf{x}^* in that

stratum when \mathbf{x}^* is restricted to avoid $\mathcal{D}_{\mathbf{x}}$. Hence, the proof of Lemma 4 in [3] of the differentiability of $\Phi(\mathbf{x}; \mathbf{x}^*)$ with respect to \mathbf{x}^* for the case that \mathbf{X}^m is a tree space and \mathbf{x}^* lies in a top-dimensional stratum will give the following generalisation of that result to the derivative of $\Phi_\sigma(\mathbf{x}; \cdot)$. Since the proof is similar to that for Lemma 4 in [3], we omit it here.

Proposition 4. *Let \mathbf{x} and \mathbf{x}^* be fixed points in \mathbf{X}^m with \mathbf{x}^* in the stratum $\sigma = \mathcal{O}(E)$ and $\mathbf{x} \notin \mathcal{D}_{\mathbf{x}^*}$. Then, the map*

$$\sigma \rightarrow \mathbb{R}(E); \quad \mathbf{x}' \mapsto \Phi_\sigma(\mathbf{x}; \mathbf{x}')$$

is differentiable with respect to \mathbf{x}' at \mathbf{x}^ with derivative given by*

$$M_{\mathbf{x}^*}^\sigma(\mathbf{x}) = J^\top \text{diag} \left\{ \mathbf{0}_{|B_0 \cap E|}, -\|B_1\|_{\mathbf{x}} M_{(A_1)_{\mathbf{x}^*}}^\dagger, \dots, -\|B_k\|_{\mathbf{x}} M_{(A_k)_{\mathbf{x}^*}}^\dagger \right\} J \quad (8)$$

where the sequences $\mathcal{A} = (A_0, \dots, A_k)$ and $\mathcal{B} = (B_0, \dots, B_k)$ form the support of the geodesic from \mathbf{x}^ to \mathbf{x} and J is the matrix representation of the map j in the expression (3) for $\Phi(\mathbf{x}; \mathbf{x}^*)$ and where, for $\mathbf{y} = (y_1, \dots, y_s) \neq 0$,*

$$M_{\mathbf{y}}^\dagger = \frac{1}{\|\mathbf{y}\|} I_s - \frac{1}{\|\mathbf{y}\|^3} \mathbf{y}^\top \mathbf{y} \quad (9)$$

is the derivative of the map $\mathbf{y} \mapsto \frac{1}{\|\mathbf{y}\|} \mathbf{y}$.

Note that, when $s = 1$, $M_{y_1}^\dagger = 0$ and, in general, that $\|\mathbf{y}\| M_{\mathbf{y}}^\dagger$ is the projection onto the hyperplane in \mathbb{R}^s orthogonal to \mathbf{y} .

Returning to $\Psi(\mathbf{x}, \mathbf{w}_\tau; \mathbf{x}^*)$ with $\mathbf{x}^* \in \sigma = \mathcal{O}(E)$, where $\tau = \mathcal{O}(E \cup F)$ co-bounds σ and \mathbf{w}_τ is in $\mathbb{R}(E) \times \mathcal{O}(F)$, since $\Psi(\mathbf{x}, \mathbf{w}_\tau; \mathbf{x}^*)$ is in \mathcal{C}_τ , both $\Psi_\tau(\mathbf{x}, \mathbf{w}_\tau; \mathbf{x}^*) = P_\tau(\Psi(\mathbf{x}, \mathbf{w}_\tau; \mathbf{x}^*))$ and $P_\sigma(\Psi(\mathbf{x}, \mathbf{w}_\tau; \mathbf{x}^*))$ are well defined. In particular, $\Psi_\tau(\cdot, \mathbf{w}_\tau; \mathbf{x}^*)$ is a map from \mathbf{X}^m onto $\mathbb{R}(E \cup F)$. Then, we also have the following consequences of Theorem 2.

Corollary 2. *With the notation and hypotheses of Theorem 2(ii),*

- (i) $\lim_{\lambda \rightarrow 0+} \Phi_\tau(\mathbf{x}; \mathbf{x}^*(\lambda, \mathbf{w}_\tau)) = \lim_{\lambda \rightarrow 0+} \Phi_\tau(\mathbf{x}; \mathbf{x}^*(\lambda, \mathbf{w}_\tau^\perp)) = \Psi_\tau(\mathbf{x}, \mathbf{w}_\tau; \mathbf{x}^*);$
- (ii) $P_\sigma(\Psi_\tau(\mathbf{x}, \mathbf{w}_\tau; \mathbf{x}^*)) = \Phi_\sigma(\mathbf{x}; \mathbf{x}^*).$

Proof. The equality of the extreme terms in (i) follows since the W_i in (7) are determined by the axes in $E \cup F$, so that it does not matter whether we project on $\mathcal{O}(E \cup F)$ before or after taking the limit, and the remaining term $(B_0 \cap (E \cup F))_{\mathbf{x}}$ remains constant throughout the limiting process. The equality with the central term in (i) follows from Corollary 1(ii): $\Psi_\tau(\mathbf{x}, \mathbf{w}_\tau; \mathbf{x}^*) = \Psi_\tau(\mathbf{x}, \mathbf{w}_\tau^\perp; \mathbf{x}^*)$, which is $\lim_{\lambda \rightarrow 0+} \Phi_\tau(\mathbf{x}; \mathbf{x}^*(\lambda, \mathbf{w}_\tau^\perp))$ by the case already established.

Note that, since projection onto $\mathbb{R}(E) \subset \mathbb{R}(E \cup F)$ is unaffected by first projecting onto $\mathbb{R}(E \cup F)$, (ii) is equivalent to

$$P_\sigma(\Psi(\mathbf{x}, \mathbf{w}_\tau; \mathbf{x}^*)) = \Phi_\sigma(\mathbf{x}; \mathbf{x}^*). \quad (10)$$

To establish (10), we need to allow for the fact that the geodesics γ_λ and the geodesic γ_0 from \mathbf{x}^* to \mathbf{x} may have different carriers. We assume that λ is

restricted to the range $0 < \lambda < \epsilon$ such that the initial segments of γ_λ all lie in $\zeta = \mathcal{O}(E \cup F \cup G)$, where possibly $G = \emptyset$ and so $\zeta = \tau$, and let K be the set of axes that are positive along an initial segment of γ_0 . Then, $K \supseteq E \cup G$. Now, $e \in E \cup F \cup G$ if, and only if, for each λ and some maximal $\delta(\lambda) > 0$, $(\gamma_\lambda(s))_e > 0$ for $s \in (0, \delta(\lambda))$. From the uniform continuity of geodesics with respect to their endpoints, it is clear that we must have $\delta(\lambda) \rightarrow \delta_0 \geq 0$ as $\lambda \rightarrow 0$. If $\delta_0 > 0$, then $(\gamma_0(s))_e > 0$ for $s \in (0, \delta_0)$ and so $e \in K$. Conversely, $e \in K \cap (E \cup F \cup G)$ implies that $(\gamma_0(s))_e > 0$ for $s \in (0, \delta(0))$ and we must have $\delta_0 = \delta(0)$.

Thus, for any axis e in $K \cap (E \cup F \cup G)$, the projections $(\gamma_\lambda(s))_e$ and $(\gamma_0(s))_e$ of the initial segments of these geodesics all lie in the closure of the stratum $\mathcal{O}(E \cup F \cup G)$. The uniform continuity of these geodesics, and so of their projections, with respect to their endpoints, together with their linearity within that closed stratum, implies that the components $(\dot{\gamma}_\lambda(0))_e$ converge to $(\dot{\gamma}_0(0))_e$ as $\lambda \rightarrow 0$. In particular, since $E \subseteq K$, this is valid for any axis e in E . This establishes (10). \square

The comments made prior to Proposition 4 regarding the algebraic expression for $\Phi_\sigma(\mathbf{x}; \mathbf{x}^*)$ can be generalised to apply to $\Psi_\tau(\mathbf{x}, \mathbf{w}_\tau; \mathbf{x}^*)$: using the notation in Theorem 2 for $\Psi(\mathbf{x}, \mathbf{w}_\tau; \mathbf{x}^*)$ we have that

$$\Psi_\tau(\mathbf{x}, \mathbf{w}_\tau; \mathbf{x}^*) = j \left((B_0 \cap (E \cup F))_{\mathbf{x}}, -\frac{\|B_1\|_{\mathbf{x}}}{\|W_1\|} W_1, \dots, -\frac{\|B_k\|_{\mathbf{x}}}{\|W_k\|} W_k \right). \quad (11)$$

Recall that $\mathcal{S}_{\tau \setminus \sigma}^{l-l'}$ denotes the set of unit vectors in $\{\mathbf{0}\} \times \mathcal{O}(F) \subset \mathbb{R}(E) \times \mathcal{O}(F)$ that comprises all unit vectors that are tangent to τ and orthogonal to σ . If $l-l' = 1$, $\mathcal{S}_{\tau \setminus \sigma}^{l-l'}$ comprises a single point. When $l-l' > 1$, for any fixed $\mathbf{x} \in \mathbf{X}^m$, the pseudo-partition of \mathbf{X}^m determined by $\mathcal{D}_{\mathbf{x}}$ induces a polyhedral subdivision of $\mathcal{S}_{\tau \setminus \sigma}^{l-l'}$ where, in each cell of the induced polyhedral subdivision, the algebraic expression (7) for $\Psi(\mathbf{x}, \cdot; \mathbf{x}^*)$, and so the algebraic expression for $\Psi_\tau(\mathbf{x}, \cdot; \mathbf{x}^*)$, remains the same. In particular, this implies that, for fixed \mathbf{x} , $\Psi_\tau(\mathbf{x}, \mathbf{w}_\tau; \mathbf{x}^*)$ is a continuous function of $\mathbf{w}_\tau \in \mathcal{S}_{\tau \setminus \sigma}^{l-l'}$. In fact, the directional derivatives of $\Psi_\tau(\mathbf{x}, \mathbf{w}_\tau; \mathbf{x}^*)$ with respect to \mathbf{w}_τ also exist in directions \mathbf{v} in the tangent space to $\mathcal{S}_{\tau \setminus \sigma}^{l-l'}$ at \mathbf{w}_τ that we denote by $\mathcal{T}_{\mathbf{w}_\tau}(\mathcal{S}_{\tau \setminus \sigma}^{l-l'})$. These derivatives have the property given in the following proposition, where we note that $\mathbb{R}(E) \times \mathcal{O}(F) \subset \mathbb{R}(E) \times \mathbb{R}(F)$ so that, for fixed \mathbf{x} and \mathbf{x}^* , \mathbf{w}_τ and $\Psi_\tau(\mathbf{x}, \mathbf{w}_\tau; \mathbf{x}^*)$ lie in the same Euclidean space.

Proposition 5. *Let the stratum $\sigma = \mathcal{O}(E)$ of co-dimension $l (\geq 2)$ bound, in \mathbf{X}^m , the stratum $\tau = \mathcal{O}(E \cup F)$ of co-dimension $l' (< l-1)$. Fix $\mathbf{x}, \mathbf{x}^* \in \mathbf{X}^m$ with $\mathbf{x}^* \in \sigma$. Then, as a function of $\mathbf{w}_\tau \in \mathcal{S}_{\tau \setminus \sigma}^{l-l'}$, the directional derivative D of $\Psi_\tau(\mathbf{x}, \mathbf{w}_\tau; \mathbf{x}^*)$ at \mathbf{w}_τ in the direction $\mathbf{v} \in \mathcal{T}_{\mathbf{w}_\tau}(\mathcal{S}_{\tau \setminus \sigma}^{l-l'})$ exists and satisfies*

$$\langle \mathbf{w}_\tau, D\Psi_\tau(\mathbf{x}, \mathbf{w}_\tau; \mathbf{x}^*)(\mathbf{v}) \rangle = 0.$$

Proof. Without loss of generality, we may assume that $\|\mathbf{v}\| = 1$. Consider the geodesic on $\mathcal{S}_{\tau \setminus \sigma}^{l-l'}$ given by $\alpha(s) = \mathbf{w}_\tau \cos s + \mathbf{v} \sin s$. Write w_1 for a vector whose coordinates comprise a subset of those of \mathbf{w}_τ , and v_1, α_1 for the corresponding

components of \mathbf{v} and α respectively. Then, the initial tangent vector of the function $f(s) = \frac{\alpha_1(s)}{\|\alpha_1(s)\|}$ is $f'(0) = v_1 M_{w_1}^\dagger$, where $M_{\mathbf{y}}^\dagger$ is given by (9). Clearly, $\langle w_1, f'(0) \rangle = 0$, since the image of $M_{w_1}^\dagger$ is orthogonal to w_1 .

On the other hand, it follows from the argument in the proof of Theorem 2 that, for all sufficiently small s , $\Psi_\tau(\mathbf{x}, \alpha(s); \mathbf{x}^*)$ all have the same algebraic expression provided that, when \mathbf{w}_τ lies on the boundary of a cell of the induced polyhedral subdivision on $\mathcal{S}_{\tau \setminus \sigma}^{l-l'}$, we use for \mathbf{w}_τ the expression valid for $s > 0$. Thus, we may use the algebraic expression for $\Psi_\tau(\mathbf{x}, \mathbf{w}_\tau; \mathbf{x}^*)$ given by (11) to express $D\Psi_\tau(\mathbf{x}, \mathbf{w}_\tau; \mathbf{x}^*)(\mathbf{v})$ in the form $\mathbf{v} M_{\mathbf{x}^*, \mathbf{x}}(\mathbf{w}_\tau)$, where

$$M_{\mathbf{x}^*, \mathbf{x}}(\mathbf{w}_\tau) = J^\top \text{diag} \left\{ \mathbf{0}, -\|B_1\|_{\mathbf{x}} M_{W_1}^\dagger, \dots, -\|B_j\|_{\mathbf{x}} M_{W_j}^\dagger, \mathbf{0} \right\} J \quad (12)$$

and where, using the notation of Theorem 2, $W_i = (A_i \cap F)_{\mathbf{w}_\tau}$, $i = 1, \dots, j$, are just those components in the expression for $\Psi_\tau(\mathbf{x}, \mathbf{w}_\tau; \mathbf{x}^*)$ for which $(A_i \cap E)_{\mathbf{x}^*} = 0$. Hence, the result follows. \square

The proof of Proposition 5 also shows that, if $\Psi_\tau(\mathbf{x}, \alpha(s); \mathbf{x}^*)$ lies in the interior of a single cell of the induced polyhedral subdivision of $\mathcal{S}_{\tau \setminus \sigma}^{l-l'}$ for all sufficiently small s , then $\Psi_\tau(\mathbf{x}, \alpha(s); \mathbf{x}^*)$ is differentiable with respect to $\mathbf{w}_\tau \in \mathcal{S}_{\tau \setminus \sigma}^{l-l'}$. However, if $\alpha(0)$ lies in the boundary of a cell of the induced polyhedral subdivision, then the algebraic expression for $\Psi_\tau(\mathbf{x}, \alpha(0); \mathbf{x}^*)$ may require to be modified by combining, or splitting, sets of components in the expression given in (7), depending on \mathbf{v} , in order that all $\Psi_\tau(\mathbf{x}, \alpha(s); \mathbf{x}^*)$ have the same algebraic expression. This in particular means that the integer k determining the number of components, in this case, depends implicitly also on \mathbf{v} , so that $\Psi_\tau(\mathbf{x}, \alpha(s); \mathbf{x}^*)$ is no longer differentiable.

The directional derivative of $\langle \mathbf{w}_\tau, \Psi_\tau(\mathbf{x}, \mathbf{w}_\tau; \mathbf{x}^*) \rangle$, as a function of \mathbf{w}_τ , now follows from Proposition 5.

Corollary 3. *Assume that all assumptions in Proposition 5 hold. Then, for any $\mathbf{v} \in \mathcal{T}_{\mathbf{w}_\tau}(\mathcal{S}_{\tau \setminus \sigma}^{l-l'})$, the derivative D in the direction \mathbf{v} of $\langle \mathbf{w}_\tau, \Psi_\tau(\mathbf{x}, \mathbf{w}_\tau; \mathbf{x}^*) \rangle$ at \mathbf{w}_τ is given by*

$$D(\langle \mathbf{w}_\tau, \Psi_\tau(\mathbf{x}, \mathbf{w}_\tau; \mathbf{x}^*) \rangle)(\mathbf{v}) = \langle \mathbf{v}, \Psi_\tau(\mathbf{x}, \mathbf{w}_\tau; \mathbf{x}^*) \rangle.$$

Proof. The second term in the expansion

$$\begin{aligned} & D(\langle \mathbf{w}_\tau, \Psi_\tau(\mathbf{x}, \mathbf{w}_\tau; \mathbf{x}^*) \rangle)(\mathbf{v}) \\ &= \langle D\mathbf{w}_\tau(\mathbf{v}), \Psi_\tau(\mathbf{x}, \mathbf{w}_\tau; \mathbf{x}^*) \rangle + \langle \mathbf{w}_\tau, D\Psi_\tau(\mathbf{x}, \mathbf{w}_\tau; \mathbf{x}^*)(\mathbf{v}) \rangle \end{aligned}$$

vanishes by Proposition 5. The result then follows since the directional derivative $D\mathbf{w}_\tau(\mathbf{v})$ is given by the derivative at $s = 0$ of the geodesic $\alpha(s) = \mathbf{w}_\tau \cos s + \mathbf{v} \sin s$. \square

5 Characterisation of Fréchet means

In the remainder of this paper, we use the knowledge obtained so far on the logarithm map to investigate Fréchet means of probability measures on \mathbf{X}^m .

So, from now on we assume that μ is a probability measure on \mathbf{X}^m and that its Fréchet function defined by (1), where $\mathbf{M} = \mathbf{X}^m$, is finite at one point. The latter ensures that the Fréchet function of μ is finite everywhere.

Since the squared distance on a CAT(0)-space is a convex function with respect to each of its variables, it follows that the Fréchet mean of μ is unique and that the condition for \mathbf{x}^* to be the Fréchet mean of μ , that is, the condition for \mathbf{x}^* to satisfy

$$\int_{\mathbf{X}^m} d(\mathbf{x}^*, \mathbf{x})^2 d\mu(\mathbf{x}) < \int_{\mathbf{X}^m} d(\mathbf{x}', \mathbf{x})^2 d\mu(\mathbf{x}), \quad \text{for any } \mathbf{x}' \neq \mathbf{x}^*,$$

is equivalent to this inequality holding in any neighbourhood of \mathbf{x}^* . Then, since the Fréchet function of μ is differentiable at \mathbf{x}^* if \mathbf{x}^* lies in a top-dimensional, or locally top-dimensional, stratum, the above condition for such \mathbf{x}^* to be the Fréchet mean of μ is equivalent to the condition that

$$\int_{\mathbf{X}^m} \log_{\mathbf{x}^*}(\mathbf{x}) d\mu(\mathbf{x}) = 0, \quad (13)$$

similar to the condition for Fréchet means in Riemannian manifolds of non-positive curvature.

When \mathbf{x}^* lies in a stratum σ of positive co-dimension that is not locally top-dimensional, the squared distance $d(\mathbf{x}^*, \mathbf{x})^2$ is no longer differentiable at \mathbf{x}^* for any fixed \mathbf{x} . Nevertheless, it has directional derivatives along all possible directions and so the above condition becomes that, at $\mathbf{x}^* \in \sigma$, the Fréchet function of μ has non-negative directional derivatives along all possible directions. The fact that \mathbf{X}^m is a CAT(0)-space also implies that the derivative at \mathbf{x}^* in the direction \mathbf{w} of the distance function $d_{\mathbf{x}} = d(\cdot, \mathbf{x})$ can be expressed as

$$(Dd_{\mathbf{x}}(\mathbf{x}^*))(\mathbf{w}) = -\frac{1}{d_{\mathbf{x}}(\mathbf{x}^*)} \ll \mathbf{w}, \log_{\mathbf{x}^*}(\mathbf{x}) \gg,$$

where $\ll \cdot, \cdot \gg$ is defined by (2) (cf. [14], (2.5), p417). Thus, the criterion for a point \mathbf{x}^* lying in a stratum σ of positive co-dimension to be the Fréchet mean of μ is equivalent to the condition that

$$\int_{\mathbf{X}^m} \ll \mathbf{w}, \log_{\mathbf{x}^*}(\mathbf{x}) \gg d\mu(\mathbf{x}) \leq 0 \quad (14)$$

for all tangent vectors \mathbf{w} at \mathbf{x}^* .

For any vector \mathbf{w} at \mathbf{x}^* which is tangent to σ , the fact that $-\mathbf{w}$ is also tangent to σ at \mathbf{x}^* implies that the inequality (14) must be an equality for all such \mathbf{w} . From this it follows that

$$\int_{\mathbf{X}^m} \log_{\mathbf{x}^*}^{\sigma}(\mathbf{x}) d\mu(\mathbf{x}) = 0, \quad (15)$$

analogous to the condition (13). On the other hand, for any given stratum τ co-bounding σ and any vector \mathbf{w} at \mathbf{x}^* tangent to τ , it is possible to link the derivative, at \mathbf{x}^* , of the Fréchet function in the direction \mathbf{w} with $\Psi_{\tau}(\cdot, \mathbf{w}; \mathbf{x}^*)$. To show this, we need the following limiting property of the directional derivatives on general CAT(0)-spaces.

Lemma 1. *Let X be a CAT(0)-space, and let x_0 and x be two distinct fixed points in X . For some $\epsilon > 0$, assume that $\gamma : [0, \epsilon) \rightarrow X$ is a geodesic with $\gamma(0) = x$ and $\gamma'(0) = v_x$. Then, if $\{x_i : i \geq 1\}$ is a sequence of points along γ convergent to x , the derivative D at x in the direction v_x of the distance function $d_{x_0} = d(x_0, \cdot)$ has the property that*

$$Dd_{x_0}(v_x) = \lim_{i \rightarrow \infty} Dd_{x_0}(v_{x_i}),$$

where v_{x_i} denotes the tangent vector at x_i of the geodesic γ .

Proof. Since $Dd_{x_0}(v_x) = -\ll v_x, \log_x(x_0) \gg / d_{x_0}(x) = -\cos \angle_x(x', x_0)$ where x' is a point on the geodesic γ , it is sufficient to show that, for a fixed point x' chosen on γ , $\angle_x(x', x_0) = \lim_{i \rightarrow \infty} \angle_{x_i}(x', x_0)$.

For this, we write $\gamma_{a,b}$ for the (unique) geodesic segment joining a and b , for any two distinct points a and b in X . Then, given sequences of points $a_i \rightarrow a$, $b_i \rightarrow b$ and $c_i \rightarrow c$ in X , it follows from the Cartan-Hadamand theorem that the geodesic segments γ_{a_i, b_i} and γ_{a_i, c_i} converge uniformly, as maps, to $\gamma_{a,b}$ and $\gamma_{a,c}$ respectively. From this it follows that $\angle_a(b, c) \geq \limsup_{i \rightarrow \infty} \angle_{a_i}(b_i, c_i)$ (cf. [7], Theorem 4.3.11, p.119). Applying this to the sequence of geodesic triangles $\Delta(x'x_ix_0)$, we obtain

$$\angle_x(x', x_0) \geq \limsup_{i \rightarrow \infty} \angle_{x_i}(x', x_0). \quad (16)$$

On the other hand, using (4.3) p.124 of [7], we have

$$\limsup_{i \rightarrow \infty} \tilde{\angle}_{x_i}(x, x_0) \leq \pi - \angle_x(x', x_0),$$

where $\tilde{\angle}$ denotes the corresponding comparison angle in \mathbb{R}^2 . Then, since $\tilde{\angle}_{x_i}(x, x_0) \geq \angle_{x_i}(x, x_0)$, the above implies that

$$\begin{aligned} \angle_x(x', x_0) &\leq \pi - \limsup_{i \rightarrow \infty} \tilde{\angle}_{x_i}(x, x_0) \\ &\leq \pi - \limsup_{i \rightarrow \infty} \angle_{x_i}(x, x_0) \\ &= \liminf_{i \rightarrow \infty} \{\pi - \angle_{x_i}(x, x_0)\}. \end{aligned}$$

However, since X has non-positive curvature, if x_i lies between x and x' on the geodesic segment $\gamma_{x,x'}$, then $\angle_{x_i}(x', x_0) + \angle_{x_i}(x_0, x) \geq \pi$ (cf. [7], p117, line 5). Hence,

$$\angle_x(x', x_0) \leq \liminf_{i \rightarrow \infty} \angle_{x_i}(x', x_0).$$

This, together with (16), gives that

$$\angle_x(x', x_0) = \lim_{i \rightarrow \infty} \angle_{x_i}(x', x_0),$$

so that the required result follows. \square

Recalling that $\Phi(\mathbf{x}; \mathbf{x}^*) = \log_{\mathbf{x}^*}(\mathbf{x}) + \mathbf{x}^*$, the criteria (14) and (15) for a point \mathbf{x}^* to be the Fréchet mean of μ may now be recast, the former in terms of the standard Euclidean inner product and the projection $\Psi_\tau(\mathbf{x}, \mathbf{w}; \mathbf{x}^*)$ when \mathbf{x}^* lies in a stratum σ of positive co-dimension and \mathbf{w} is tangent to a co-bounding stratum τ .

Theorem 3. *Let σ be a stratum in \mathbf{X}^m of co-dimension $l(\geq 0)$. The necessary and sufficient conditions for a given point $\mathbf{x}^* \in \sigma$ to be the Fréchet mean of μ are*

- (i) *for any stratum τ in \mathbf{X}^m of co-dimension l' , $0 \leq l' < l$, co-bounding σ and any $\mathbf{w}_\tau \in \mathcal{S}_{\tau \setminus \sigma}^{l-l'}$,*

$$\left\langle \mathbf{w}_\tau, \int_{\mathbf{X}^m} \Psi_\tau(\mathbf{x}, \mathbf{w}_\tau; \mathbf{x}^*) d\mu(\mathbf{x}) \right\rangle \leq 0; \quad (17)$$

- (ii) *for all $l \geq 0$,*

$$\mathbf{x}^* = \int_{\mathbf{X}^m} \Phi_\sigma(\mathbf{x}; \mathbf{x}^*) d\mu(\mathbf{x}). \quad (18)$$

Note that case (i) may only occur if $l > 0$, but need not occur then. Note also that, if \mathbf{X}^m is a tree space, the special case $l = 0$ of this result is the same as that of Lemma 3 of [3]; and the special case $l = 1$, so that $l' = 0$, is equivalent to that given by Lemma 5 of [3]: on the one hand, $\mathcal{S}_{\tau \setminus \sigma}^{l-l'}$ contains a single unit vector and, on the other hand, as we noted earlier, $\Psi_\tau(\cdot, \mathbf{w}_\tau; \mathbf{x}^*) = \Psi(\cdot, \mathbf{w}_\tau; \mathbf{x}^*)$ is identical with the composition of the ‘folding map’ with $\Phi(\cdot; \mathbf{x}^*)$ in [3].

Proof. Noting that (ii) is precisely (15), it is sufficient to show that (i) is equivalent to (14) for any tangent vector \mathbf{w} that is not tangent to $\sigma = \mathcal{O}(E)$. For this, we fix any stratum $\tau = \mathcal{O}(E \cup F)$, of co-dimension l' , co-bounding $\sigma = \mathcal{O}(E)$ and take $\mathbf{w} = \mathbf{w}_\tau \in \mathbb{R}(E) \times \mathcal{O}(F)$. Then, it follows from Lemma 1 that (14) is equivalent to

$$\lim_{\lambda \rightarrow 0+} \int_{\mathbf{X}^m} \ll \mathbf{w}_\tau, \log_{\mathbf{x}^*(\lambda, \mathbf{w}_\tau)}(\mathbf{x}) \gg d\mu(\mathbf{x}) \leq 0, \quad (19)$$

where $\mathbf{x}^*(\lambda, \mathbf{w}_\tau)$ is as defined prior to Theorem 2. Since \mathbf{w}_τ is tangent to τ at $\mathbf{x}^*(\lambda, \mathbf{w}_\tau)$ for sufficiently small $\lambda > 0$ and, for any given \mathbf{x} , $\log_{\mathbf{x}^*(\lambda, \mathbf{w}_\tau)}(\mathbf{x})$ is tangent either to τ or to one of the strata that co-bound τ , we have

$$\ll \mathbf{w}_\tau, \log_{\mathbf{x}^*(\lambda, \mathbf{w}_\tau)}(\mathbf{x}) \gg = \langle \mathbf{w}_\tau, \log_{\mathbf{x}^*(\lambda, \mathbf{w}_\tau)}(\mathbf{x}) \rangle.$$

However,

$$\begin{aligned} \left\langle \mathbf{w}_\tau, \log_{\mathbf{x}^*(\lambda, \mathbf{w}_\tau)}(\mathbf{x}) \right\rangle &= \langle \mathbf{w}_\tau, \Phi(\mathbf{x}; \mathbf{x}^*(\lambda, \mathbf{w}_\tau)) - \mathbf{x}^*(\lambda, \mathbf{w}_\tau) \rangle \\ &= \langle \mathbf{w}_\tau, \Phi_\tau(\mathbf{x}; \mathbf{x}^*(\lambda, \mathbf{w}_\tau)) - \mathbf{x}^*(\lambda, \mathbf{w}_\tau) \rangle. \end{aligned}$$

Hence, by Corollary 2(i) and then Corollary 1(ii), (19) is equivalent to

$$\int_{\mathbf{X}^m} \langle \mathbf{w}_\tau, \Psi_\tau(\mathbf{x}, \mathbf{w}_\tau^\perp; \mathbf{x}^*) \rangle d\mu(\mathbf{x}) \leq \langle \mathbf{w}_\tau, \mathbf{x}^* \rangle,$$

where $\mathbf{w}_\tau^\perp = P_{\tau \setminus \sigma}(\mathbf{w}_\tau)$. Decomposing \mathbf{w}_τ as $\mathbf{w}_\tau = \mathbf{w}_\sigma + \mathbf{w}_\tau^\perp$, where $\mathbf{w}_\sigma = P_\sigma(\mathbf{w}_\tau)$, leads to

$$\begin{aligned} \langle \mathbf{w}_\tau, \Psi_\tau(\mathbf{x}, \mathbf{w}_\tau^\perp; \mathbf{x}^*) \rangle &= \langle \mathbf{w}_\sigma, \Psi_\tau(\mathbf{x}, \mathbf{w}_\tau^\perp; \mathbf{x}^*) \rangle + \langle \mathbf{w}_\tau^\perp, \Psi_\tau(\mathbf{x}, \mathbf{w}_\tau^\perp; \mathbf{x}^*) \rangle \\ &= \langle \mathbf{w}_\sigma, \Phi_\sigma(\mathbf{x}; \mathbf{x}^*) \rangle + \langle \mathbf{w}_\tau^\perp, \Psi_\tau(\mathbf{x}, \mathbf{w}_\tau^\perp; \mathbf{x}^*) \rangle, \end{aligned}$$

where the second equality follows from Corollary 2(ii). The required result now follows by noting (ii), noting that $\langle \mathbf{w}_\tau, \mathbf{x}^* \rangle = \langle \mathbf{w}_\sigma, \mathbf{x}^* \rangle = 0$ and noting that, by applying the projection P_τ to the result of Corollary 1(i), $\Psi_\tau(\mathbf{x}, \mathbf{w}_\tau^\perp; \mathbf{x}^*) = \Psi_\tau(\mathbf{x}, \mathbf{w}_\tau^\perp / \|\mathbf{w}_\tau^\perp\|; \mathbf{x}^*)$. \square

When the stratum containing the Fréchet mean \mathbf{x}^* of the probability measure μ on \mathbf{X}^m is of positive co-dimension, (17) being an equality has a significant influence on the nature of the distributions of the Euclidean random variables $\Psi_\tau(\xi, \mathbf{w}_\tau; \mathbf{x}^*)$, where ξ is a random variable on \mathbf{X}^m with distribution μ .

Definition 10. For the stratum σ of co-dimension $l(\geq 1)$, in which the Fréchet mean \mathbf{x}^* of μ lies, and the stratum τ , of co-dimension l' , co-bounding σ , the subset $\Theta_{\tau,\sigma}(\mathbf{x}^*; \mu)$ of $\mathcal{S}_{\tau \setminus \sigma}^{l-l'}$ is defined as

$$\Theta_{\tau,\sigma}(\mathbf{x}^*; \mu) = \left\{ \mathbf{w}_\tau \in \mathcal{S}_{\tau \setminus \sigma}^{l-l'} \mid \text{the inequality (17) for } \mathbf{w}_\tau \text{ is an equality} \right\}. \quad (20)$$

The convexity of the directional derivative $D(d_{\mathbf{x}}^2)(\mathbf{w})$ in \mathbf{w} (cf. [14], pp416-417) ensures that $\Theta_{\tau,\sigma}(\mathbf{x}^*; \mu)$ is a convex subset of $\mathcal{S}_{\tau \setminus \sigma}^{l-l'}$ and that

$$\Theta_\sigma(\mathbf{x}^*; \mu) = \bigcup_{\tau \supset \sigma} \Theta_{\tau,\sigma}(\mathbf{x}^*; \mu) \quad (21)$$

is a convex subset of $\bigcup_{\tau \supset \sigma} \mathcal{S}_{\tau \setminus \sigma}^{l-l'} \subseteq \mathcal{C}_\sigma$. If $l-l' = 1$, $\mathcal{S}_{\tau \setminus \sigma}^{l-l'}$ consists of a single unit vector so that $\Theta_{\tau,\sigma}(\mathbf{x}^*; \mu)$ is either $\mathcal{S}_{\tau \setminus \sigma}^{l-l'}$ itself or an empty set. In general, if the closure of $\Theta_{\tau,\sigma}(\mathbf{x}^*; \mu)$ is contained in $\mathcal{S}_{\tau \setminus \sigma}^{l-l'}$, the fact that $\langle \mathbf{w}_\tau, \Psi_\tau(\mathbf{x}, \mathbf{w}_\tau; \mathbf{x}^*) \rangle$ is continuous in $\mathbf{w}_\tau \in \mathcal{S}_{\tau \setminus \sigma}^{l-l'}$ implies that $\Theta_{\tau,\sigma}(\mathbf{x}^*; \mu)$ itself must be closed.

The following result gives a relationship between the Fréchet mean \mathbf{x}^* of μ and the Euclidean mean of $\Psi_\tau(\xi, \mathbf{w}_\tau; \mathbf{x}^*)$. Here, and henceforth, by interior we intend the relative interior, that is, interior with respect to the subspace topology.

Proposition 6. Let the stratum σ of co-dimension $l(\geq 2)$ bound, in \mathbf{X}^m , the stratum τ of co-dimension $l'(< l-1)$. Assume that the Fréchet mean \mathbf{x}^* of μ lies in σ and that $\text{int}(\Theta_{\tau,\sigma}(\mathbf{x}^*; \mu)) \neq \emptyset$. Then, for any $\mathbf{w}_\tau \in \Theta_{\tau,\sigma}(\mathbf{x}^*; \mu)$,

$$\int_{\mathbf{X}^m} \{ \Psi_\tau(\mathbf{x}, \mathbf{w}_\tau; \mathbf{x}^*) - \Phi_\sigma(\mathbf{x}; \mathbf{x}^*) \} d\mu(\mathbf{x}) = 0. \quad (22)$$

Note that, if $l' = l-1$, equality (22) holds automatically since its left hand side is a 1-dimensional vector so that the equality follows from the assumption that $\mathbf{w}_\tau \in \Theta_{\tau,\sigma}(\mathbf{x}^*; \mu)$.

Proof. By the continuity of $\Psi_\tau(\mathbf{x}, \mathbf{w}_\tau; \mathbf{x}^*)$ in \mathbf{w}_τ , we may assume that $\mathbf{w}_\tau \in \text{int}(\Theta_{\tau,\sigma}(\mathbf{x}^*; \mu))$. Then equality holds in (17) in a neighbourhood of \mathbf{w}_τ , so that

$$D \left(\left\langle \mathbf{w}_\tau, \int_{\mathbf{X}^m} \Psi_\tau(\mathbf{x}, \mathbf{w}_\tau; \mathbf{x}^*) d\mu(\mathbf{x}) \right\rangle \right) (\mathbf{v}) = 0, \quad \forall \mathbf{v} \in \mathcal{T}_{\mathbf{w}_\tau}(\mathcal{S}_{\tau \setminus \sigma}^{l-l'}).$$

By Corollary 3, this implies that

$$\left\langle \mathbf{v}, \int_{\mathbf{X}^m} \Psi_\tau(\mathbf{x}, \mathbf{w}_\tau; \mathbf{x}^*) d\mu(\mathbf{x}) \right\rangle = 0, \quad \forall \mathbf{v} \in \mathcal{T}_{\mathbf{w}_\tau}(\mathcal{S}_{\tau \setminus \sigma}^{l-l'}).$$

On the other hand, it follows from $\int_{\mathbf{X}^m} \Phi_\sigma(\mathbf{x}; \mathbf{x}^*) d\mu(\mathbf{x}) = \mathbf{x}^*$ and $\langle \mathbf{w}_\tau, \mathbf{x}^* \rangle = 0$ that

$$\left\langle \mathbf{w}_\tau, \int_{\mathbf{X}^m} \Phi_\sigma(\mathbf{x}; \mathbf{x}^*) d\mu(\mathbf{x}) \right\rangle = 0, \quad \forall \mathbf{w}_\tau \in \mathcal{S}_{\tau \setminus \sigma}^{l-l'}.$$

Hence, taking the directional derivative of the left hand side as a function of $\mathbf{w}_\tau \in \mathcal{S}_{\tau \setminus \sigma}^{l-l'}$, we have

$$\left\langle \mathbf{v}, \int_{\mathbf{X}^m} \Phi_\sigma(\mathbf{x}; \mathbf{x}^*) d\mu(\mathbf{x}) \right\rangle = 0$$

for all $\mathbf{v} \in \mathcal{T}_{\mathbf{w}_\tau}(\mathcal{S}_{\tau \setminus \sigma}^{l-l'})$. Noting that the left hand side of (22) is a vector lying in the $(l-l')$ -dimensional Euclidean space containing $\mathcal{S}_{\tau \setminus \sigma}^{l-l'}$, the fact that $\mathbf{w}_\tau \in \Theta_{\tau, \sigma}(\mathbf{x}^*; \mu)$, together with the above, implies that the required result holds for any $\mathbf{w}_\tau \in \text{int}(\Theta_{\tau, \sigma}(\mathbf{x}^*; \mu))$. \square

One immediate consequence of Proposition 6 is the following.

Corollary 4. *Assume that the conditions given in Proposition 6 are satisfied. If $\sigma = \mathcal{O}(E)$ and $\tau = \mathcal{O}(E \cup F)$ then, for all $\mathbf{w}_\tau \in \Theta_{\tau, \sigma}(\mathbf{x}^*; \mu)$,*

$$\mathbf{x}^* = \int_{\mathbf{X}^m} \Psi_\tau(\mathbf{x}, \mathbf{w}_\tau; \mathbf{x}^*) d\mu(\mathbf{x}). \quad (23)$$

That is, the point $\mathbf{x}^ \in \sigma$, as a point in $\mathbb{R}(E \cup F)$, is the Euclidean mean of each of the Euclidean random variables $\Psi_\tau(\cdot, \mathbf{w}_\tau; \mathbf{x}^*)$ for such \mathbf{w}_τ .*

For further investigation of the role that the set $\Theta_{\tau, \sigma}(\mathbf{x}^*; \mu)$ defined by (20) plays, we introduce another set, related to the limits of the logarithm map at \mathbf{x}^* and defined as follows.

Definition 11. *Let the stratum $\sigma = \mathcal{O}(E)$ of co-dimension $l(\geq 1)$ bound, in \mathbf{X}^m , the stratum τ of co-dimension $l'(< l)$. For $\mathbf{x}^* \in \sigma$ and $\mathbf{w}_\tau \in \mathcal{S}_{\tau \setminus \sigma}^{l-l'}$, a point $\mathbf{x} \in \mathbf{X}^m$ is called singular with respect to $(\mathbf{x}^*, \mathbf{w}_\tau)$, if at least one A_i , $i \geq 1$, with $A_i \cap E = \emptyset$ has $|(A_i)_{\mathbf{w}_\tau}| > 1$, where the sets of axes A_i are those specified in the statement of Theorem 2 for the expression of $\Psi(\mathbf{x}, \mathbf{w}_\tau; \mathbf{x}^*)$. The set $\Sigma_{\tau, \sigma}(\mathbf{x}^*; \mathbf{w}_\tau)$ consists of all points \mathbf{x} that are singular with respect to $(\mathbf{x}^*, \mathbf{w}_\tau)$.*

It follows from comparison of the corresponding expressions (7) and (11) that the singularity of \mathbf{x} with respect to $(\mathbf{x}^*, \mathbf{w}_\tau)$ has the same effect on $\Psi_\tau(\mathbf{x}, \mathbf{w}_\tau; \mathbf{x}^*)$ as it does on $\Psi(\mathbf{x}, \mathbf{w}_\tau; \mathbf{x}^*)$.

Note that $\Sigma_{\tau, \sigma}(\mathbf{x}^*; \mathbf{w}_\tau) = \emptyset$ if $l - l' = 1$, since then $\mathcal{S}_{\tau \setminus \sigma}^{l-l'}$ contains a single unit vector \mathbf{w}_τ which leads to the impossibility that $|(A_i)_{\mathbf{w}_\tau}| > 1$. Generally, if $l - l' > 1$, which implies that $l \geq 2$, $\Sigma_{\tau, \sigma}(\mathbf{x}^*; \mathbf{w}_\tau)$ could be relatively substantial. Nevertheless, we have the following result on the measure of $\Sigma_{\tau, \sigma}(\mathbf{x}^*; \mathbf{w}_\tau)$ for $\mathbf{w}_\tau \in \text{int}(\Theta_{\tau, \sigma}(\mathbf{x}^*; \mu))$.

Proposition 7. *Let the stratum σ of co-dimension $l(\geq 2)$ bound, in \mathbf{X}^m , the stratum τ of co-dimension $l'(< l - 1)$. Assume that the Fréchet mean \mathbf{x}^* of μ lies in σ and that $\mathbf{w}_\tau \in \text{int}(\Theta_{\tau,\sigma}(\mathbf{x}^*; \mu))$, where $\Theta_{\tau,\sigma}(\mathbf{x}^*; \mu)$ is defined in (20). Then, $\mu(\Sigma_{\tau,\sigma}(\mathbf{x}^*; \mathbf{w}_\tau)) = 0$.*

Proof. Let $\alpha(s)$ be a unit speed geodesic in $\mathcal{S}_{\tau \setminus \sigma}^{l-l'}$, write $\mathbf{v}(s) = \alpha'(s)$ and define $h(s) = \langle \mathbf{v}(s), \int_{\mathbf{X}^m} \Psi_\tau(\mathbf{x}, \alpha(s); \mathbf{x}^*) d\mu(\mathbf{x}) \rangle$. Since $\mathcal{S}_{\tau \setminus \sigma}^{l-l'}$ is an open subset of a Euclidean sphere, we have $\mathbf{v}'(s) = -\alpha(s)$, $\alpha''(s) = -\alpha(s)$ and so, by Proposition 5 and its proof,

$$\begin{aligned} h'(s) &= \left\langle \mathbf{v}(s), \int_{\mathbf{X}^m} D\Psi_\tau(\mathbf{x}, \alpha(s); \mathbf{x}^*)(\mathbf{v}(s)) d\mu(\mathbf{x}) \right\rangle \\ &= \left\langle \mathbf{v}(s), \int_{\Sigma_{\tau,\sigma}(\mathbf{x}^*; \alpha(s))} \mathbf{v}(s) M_{\mathbf{x}^*, \mathbf{x}}(\alpha(s)) d\mu(\mathbf{x}) \right\rangle, \end{aligned}$$

where $M_{\mathbf{x}^*, \mathbf{x}}(\mathbf{w})$ is given by (12). The expression for $M_{\mathbf{x}^*, \mathbf{x}}(\mathbf{w})$ implies that, for $\mathbf{w} \in \mathcal{S}_{\tau \setminus \sigma}^{l-l'}$ and any fixed $\mathbf{x} \in \Sigma_{\tau,\sigma}(\mathbf{x}^*; \mathbf{w})$, $\langle \mathbf{v}, \mathbf{v} M_{\mathbf{x}^*, \mathbf{x}}(\mathbf{w}) \rangle$ can be written in the form

$$-\sum_{i=1}^j \frac{\|B_i\|}{\|W_i\|} \left\{ \|v_i\|^2 - \left\langle v_i, \frac{W_i}{\|W_i\|} \right\rangle^2 \right\},$$

where W_i and B_i are those required for the expression (12) for $M_{\mathbf{x}^*, \mathbf{x}}(\mathbf{w})$ in the proof of Proposition 5. This implies that $h'(0)$ must be non-positive. Moreover, for any open or closed subset $\mathcal{E} \subseteq \Sigma_{\tau,\sigma}(\mathbf{x}^*; \alpha(0))$ such that $\Psi_\tau(\mathbf{x}, \alpha(0); \mathbf{x}^*)$ has the same expression for all $\mathbf{x} \in \mathcal{E}$, there is a vector $\mathbf{v}(0) \in \mathcal{T}_{\alpha(0)}(\mathcal{S}_{\tau \setminus \sigma}^{l-l'})$ such that $\langle \mathbf{v}(0), \mathbf{v}(0) M_{\mathbf{x}^*, \mathbf{x}}(\alpha(0)) \rangle < 0$ for all $\mathbf{x} \in \mathcal{E}$. Then, if $\mu(\mathcal{E}) \neq 0$, the corresponding h satisfies

$$h'(0) \leq \left\langle \mathbf{v}(0), \int_{\mathcal{E}} \mathbf{v}(0) M_{\mathbf{x}^*, \mathbf{x}}(\alpha(0)) d\mu(\mathbf{x}) \right\rangle < 0.$$

Clearly, $\Sigma_{\tau,\sigma}(\mathbf{x}^*; \alpha(0))$ can be decomposed as a finite disjoint union of such sets \mathcal{E} .

If $\mathbf{w}_\tau = \alpha(0) \in \text{int}(\Theta_{\tau,\sigma}(\mathbf{x}^*; \mu))$ then, for any $\mathbf{v}(0) \in \mathcal{T}_{\mathbf{w}_\tau}(\mathcal{S}_{\tau \setminus \sigma}^{l-l'})$, the corresponding geodesic $\alpha(s)$ lies in $\Theta_{\tau,\sigma}(\mathbf{x}^*; \mu)$ for all sufficiently small $s \geq 0$. Using a similar argument to that for the proof of Proposition 6, the corresponding $h(s)$ must be identically zero for all sufficiently small $s \geq 0$, which implies that $h'(0) = 0$. Hence, we must have $\mu(\Sigma_{\tau,\sigma}(\mathbf{x}^*; \mathbf{w}_\tau)) = 0$. \square

If a stratum σ bounds τ in \mathbf{X}^m , $\mathbf{x}^* \in \sigma$ and $\mathbf{w}_\tau^1, \mathbf{w}_\tau^2$ are two different vectors at \mathbf{x}^* tangent to τ , then the distribution of the Euclidean random variable $\Psi_\tau(\boldsymbol{\xi}, \mathbf{w}_\tau^1; \mathbf{x}^*)$ generally differs from that of $\Psi_\tau(\boldsymbol{\xi}, \mathbf{w}_\tau^2; \mathbf{x}^*)$. Nevertheless, under the conditions in Proposition 7, the $\Psi_\tau(\boldsymbol{\xi}, \mathbf{w}_\tau; \mathbf{x}^*)$ are in fact a.s. identical for $\mathbf{w}_\tau \in \text{int}(\Theta_{\tau,\sigma}(\mathbf{x}^*; \mu))$.

Proposition 8. *Assume that $\boldsymbol{\xi}$ is a random variable on \mathbf{X}^m with distribution μ having Fréchet mean \mathbf{x}^* . Assume further that $\mu(\mathcal{D}_{\mathbf{x}^*}) = 0$ and that \mathbf{x}^* lies in the stratum $\sigma = \mathcal{O}(E)$ of co-dimension $l(\geq 2)$. Let the stratum τ of co-dimension $l'(< l - 1)$ co-bound σ , in \mathbf{X}^m . If $\text{int}(\Theta_{\tau,\sigma}(\mathbf{x}^*; \mu)) \neq \emptyset$ then, the distributions of the Euclidean random variables $\Psi_\tau(\boldsymbol{\xi}, \mathbf{w}_\tau; \mathbf{x}^*)$ are independent of $\mathbf{w}_\tau \in \text{int}(\Theta_{\tau,\sigma}(\mathbf{x}^*; \mu))$.*

Note that the example in the next section makes it clear that the condition $\mathbf{w}_\tau \in \text{int}(\Theta_{\tau,\sigma}(\mathbf{x}^*; \mu))$ in the statement of Proposition 8 cannot be relaxed to $\mathbf{w}_\tau \in \Theta_{\tau,\sigma}(\mathbf{x}^*; \mu)$.

Proof. First, we show that, for any given distinct $\mathbf{w}_\tau^i \in S_{\tau \setminus \sigma}^{l-l'}$, $i = 1, 2$, and for $\mathbf{x} \notin \Sigma_{\tau,\sigma}(\mathbf{x}^*; \mathbf{w}_\tau^1) \cup \Sigma_{\tau,\sigma}(\mathbf{x}^*; \mathbf{w}_\tau^2) \cup \mathcal{D}_{\mathbf{x}^*}$, $\Psi_\tau(\mathbf{x}, \mathbf{w}_\tau^1; \mathbf{x}^*) = \Psi_\tau(\mathbf{x}, \mathbf{w}_\tau^2; \mathbf{x}^*)$. Then, it follows from the assumption and Proposition 7 that $\Psi_\tau(\xi, \mathbf{w}_\tau^1; \mathbf{x}^*) = \Psi_\tau(\xi, \mathbf{w}_\tau^2; \mathbf{x}^*)$ a.s. Recall that, for fixed $\mathbf{x} \in \mathbf{X}^m$, $\mathbf{x}^* \in \sigma$ and $\mathbf{w}_\tau \in S_{\tau \setminus \sigma}^{l-l'}$, the supports of the geodesics from $\mathbf{x}^*(\lambda, \mathbf{w}_\tau) = \mathbf{x}^* + \lambda \mathbf{w}_\tau$ to \mathbf{x} are the same, for all sufficiently small $\lambda > 0$, and that the expression for $\Psi_\tau(\mathbf{x}, \mathbf{w}_\tau; \mathbf{x}^*)$ is determined by this common support. Thus, $\Psi_\tau(\mathbf{x}, \mathbf{w}_\tau^i; \mathbf{x}^*)$ will certainly be identical if the geodesics from $\mathbf{x}^*(\lambda, \mathbf{w}_\tau^i)$ to \mathbf{x} have the same support when $\lambda > 0$ is sufficiently small.

Suppose now that the supports $(\mathcal{A}^i, \mathcal{B}^i)$, $i = 1, 2$, of the geodesics from $\mathbf{x}^*(\lambda, \mathbf{w}_\tau^1)$ and $\mathbf{x}^*(\lambda, \mathbf{w}_\tau^2)$ respectively to \mathbf{x} are different, for all sufficiently small $\lambda > 0$. Then, the geodesic γ_λ between $\mathbf{x}^*(\lambda, \mathbf{w}_\tau^1)$ and $\mathbf{x}^*(\lambda, \mathbf{w}_\tau^2)$ must meet at least one hyper-surface in $\mathcal{D}_{\mathbf{x}}$. If there are more than one, but necessarily finitely many, such hyper-surfaces, by introducing a point on γ_λ between each pair of consecutive such hyper-surfaces, the change of the supports of the geodesics from points of γ_λ to \mathbf{x} can be considered inductively to reduce the case to where γ_λ meets only one such hyper-surface.

Hence, without loss of generality, we assume that γ_λ only meets $\mathcal{D}_{\mathbf{x}}$ at a point p_λ of the hyper-surface H in $\mathcal{D}_{\mathbf{x}}$, where H is determined by (6) with $i = j$ and (5) with $i \neq j$, \mathbf{x}^* being replaced by p_λ in both equations. If $\mathbf{x} \notin \mathcal{D}_{\mathbf{x}^*}$ so that $\mathbf{x}^* \notin \mathcal{D}_{\mathbf{x}}$, the points $\mathbf{x}^*(\lambda, \mathbf{w}_\tau^i)$ lie on the opposite sides of H for all sufficiently small $\lambda > 0$. As γ_λ moves through p_λ , the supports of the geodesics from γ_λ to \mathbf{x} change, with two of the relevant subsets, without loss of generality A_j^1 and A_{j+1}^1 of the component \mathcal{A}^1 of the support $(\mathcal{A}^1, \mathcal{B}^1)$ on the one side, combining into one subset on the other. We show now that neither of these subsets A_j^1 and A_{j+1}^1 can meet E . If $A_j^1 \cap E = \emptyset$ but $A_{j+1}^1 \cap E \neq \emptyset$, then $(A_j^1)_{\mathbf{x}^*(\lambda, \mathbf{w}_\tau^1)} \rightarrow 0$ as $\lambda \rightarrow 0$. So, for sufficiently small $\lambda > 0$, the inequality (5) will hold with $i = j$ and \mathbf{x}^* replaced by $\gamma_\lambda(t)$ for any t such that $\gamma_\lambda(t)$ lies between $\mathbf{x}^*(\lambda, \mathbf{w}_\tau^1)$ and $\mathbf{x}^*(\lambda, \mathbf{w}_\tau^2)$. Hence, the amalgamation of A_j^1 and A_{j+1}^1 could not have occurred. Similarly, if $A_l^1 \cap E \neq \emptyset$ for $l = j, j+1$ then, as established in the proof of Corollary 2(ii), $(A_l^1)_{\gamma_\lambda(t)} \rightarrow (A_l^1)_{\mathbf{x}^*}$ as $\lambda \rightarrow 0$. Since $\mathbf{x}^* \notin \mathcal{D}_{\mathbf{x}}$, for sufficiently small $\lambda > 0$, the inequality (5) will hold with $i = j$ and \mathbf{x}^* replaced by $\gamma_\lambda(t)$ for any t , and we have a contradiction again. Thus, in the case when the supports $(\mathcal{A}^i, \mathcal{B}^i)$ are different, we still have $A_i^1 = A_i^2$ for all i such that $A_i^1 \cap E \neq \emptyset$.

If further $\mathbf{x} \notin \Sigma_{\tau,\sigma}(\mathbf{x}^*; \mathbf{w}_\tau^1) \cup \Sigma_{\tau,\sigma}(\mathbf{x}^*; \mathbf{w}_\tau^2)$ then, since $A_0^1 = A_0^2$, the above conclusion implies that the corresponding \mathcal{A}^i can only differ by a permutation of the A_j where $A_j^1 \cap E = \emptyset$. Hence, by Proposition 2, $(\mathcal{A}^i, \mathcal{B}^i)$ are permutations of each other and so, for such \mathbf{x} , $\Psi_\tau(\mathbf{x}, \mathbf{w}_\tau^2; \mathbf{x}^*) = \Psi_\tau(\mathbf{x}, \mathbf{w}_\tau^1; \mathbf{x}^*)$.

Next, assume that the two \mathbf{w}_τ^i are chosen to be sufficiently close that, for any given \mathbf{x} and all sufficiently small $\lambda > 0$, the geodesics from $\mathbf{x}^*(\lambda, \mathbf{w}_\tau^i)$ to \mathbf{x} have the same support. Then, if $\mathbf{w}_\tau(\alpha)$, $\alpha \in [0, 1]$, is the geodesic between \mathbf{w}_τ^1 and \mathbf{w}_τ^2 , an elementary argument on the relevant parameters in the inequalities (5) that determine the carrier will show that these parameters are monotonic in α along the geodesic. So, the geodesic from $\mathbf{x}^*(\lambda, \mathbf{w}_\tau(\alpha))$ to \mathbf{x} will have the

same support as that for the geodesics from $\mathbf{x}^*(\lambda, \mathbf{w}_\tau^i)$ to \mathbf{x} . This implies that $\Sigma_{\tau,\sigma}(\mathbf{x}^*; \mathbf{w}_\tau(\alpha)) \subseteq \Sigma_{\tau,\sigma}(\mathbf{x}^*; \mathbf{w}_\tau^1) \cup \Sigma_{\tau,\sigma}(\mathbf{x}^*; \mathbf{w}_\tau^2)$, so that $\Psi_\tau(\boldsymbol{\xi}, \mathbf{w}_\tau(\alpha); \mathbf{x}^*)$ are a.s. independent of $\alpha \in [0, 1]$.

Finally, since $\Theta_{\tau,\sigma}(\mathbf{x}^*; \mu)$ is convex, there is a sequence $\{\mathbf{w}_\tau^n \mid n \geq 1\} \subset \text{int}(\Theta_{\tau,\sigma}(\mathbf{x}^*; \mu))$ such that

$$\text{int}(\Theta_{\tau,\sigma}(\mathbf{x}^*; \mu)) = \lim_{n \rightarrow \infty} C_n, \quad (24)$$

where C_n is the convex hull in $\mathcal{S}_{\tau \setminus \sigma}^{l-l'}$ of $\{\mathbf{w}_\tau^1, \dots, \mathbf{w}_\tau^n\}$. The above argument implies that, without loss of generality, we may also assume that $\{\mathbf{w}_\tau^n \mid n \geq 1\}$ have the property that, for any $\mathbf{w}_\tau \in C_n$,

$$\Sigma_{\tau,\sigma}(\mathbf{x}^*; \mathbf{w}_\tau) \subseteq \bigcup_{i=1}^n \Sigma_{\tau,\sigma}(\mathbf{x}^*; \mathbf{w}_\tau^i).$$

This shows that

$$\mu \left(\bigcup_{\mathbf{w}_\tau \in C_n} \Sigma_{\tau,\sigma}(\mathbf{x}^*; \mathbf{w}_\tau) \right) = \mu \left(\bigcup_{i=1}^n \Sigma_{\tau,\sigma}(\mathbf{x}^*; \mathbf{w}_\tau^i) \right) = 0,$$

so that $\Psi_\tau(\boldsymbol{\xi}, \mathbf{w}_\tau; \mathbf{x}^*)$ are a.s. independent of $\mathbf{w}_\tau \in C_n$. Hence, it follows from (24) that

$$\mu \left(\bigcup_{\mathbf{w}_\tau \in \text{int}(\Theta_{\tau,\sigma}(\mathbf{x}^*; \mu))} \Sigma_{\tau,\sigma}(\mathbf{x}^*; \mathbf{w}_\tau) \right) = \lim_{n \rightarrow \infty} \mu \left(\bigcup_{\mathbf{w}_\tau \in C_n} \Sigma_{\tau,\sigma}(\mathbf{x}^*; \mathbf{w}_\tau) \right) = 0,$$

which gives the required result. \square

6 The limiting distribution of sample Fréchet means

Suppose that $\{\boldsymbol{\xi}_i : i \geq 1\}$ is a sequence of *i.i.d.* random variables in \mathbf{X}^m with probability measure μ and that $\hat{\boldsymbol{\xi}}_n$ is the sample Fréchet mean of $\boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_n$. Then, $\hat{\boldsymbol{\xi}}_n$ converges to the Fréchet mean \mathbf{x}^* of μ almost surely as n tends to infinity (cf. [22]).

If \mathbf{x}^* lies in a top-dimensional stratum, \mathbf{X}^m is locally an m -dimensional manifold. One would expect that the limiting behaviour of sample Fréchet means $\hat{\boldsymbol{\xi}}_n$ is similar, to some extent, to that of sample Fréchet means in a Riemannian manifold as obtained in [4] and [13]. In particular, the support of the limiting distribution of $\sqrt{n} \log_{\mathbf{x}^*}(\hat{\boldsymbol{\xi}}_n) = \sqrt{n}(\hat{\boldsymbol{\xi}}_n - \mathbf{x}^*)$ is the entire tangent space to \mathbf{X}^m at \mathbf{x}^* , as long as $\text{cov}(\Phi(\boldsymbol{\xi}_1; \mathbf{x}^*))$ has rank m . This fact was proved for the case of open books in [10] and for the case of tree spaces in [2] and [3]. We shall see in the following that the argument used in [3] can be generalised to \mathbf{X}^m , so that the corresponding conclusion is also valid for orthant spaces.

However, when \mathbf{x}^* lies in a stratum of positive co-dimension that is not locally top-dimensional, the limiting behaviour of sample Fréchet means is generally very different. In the case that \mathbf{X}^m is an open book or a tree space and

that the stratum containing \mathbf{x}^* is of the co-dimension one, this phenomenon was observed and studied in [10], [2] and [3]. Similarly, for general orthant spaces, the strictness or otherwise of the inequality (17) affects the limiting behaviour of $\hat{\xi}_n$. In particular, when (17) is strict, there is a constraint on the support of the limiting distribution. To describe this, we recall that, for $\sigma = \mathcal{O}(E)$ of co-dimension l and $\tau = \mathcal{O}(E \cup F)$ of co-dimension $l' < l$ co-bounding σ , we are denoting the set of unit vectors in $\mathbb{R}(E) \times \mathcal{O}(F)$ by $\mathcal{S}_{\tau, \sigma}^{m-l'}$ and those in $\{\mathbf{0}\} \times \mathcal{O}(F)$ by $\mathcal{S}_{\tau \setminus \sigma}^{l-l'}$. Then, for \mathbf{w}_τ in the latter, denote by $\mathcal{H}_{\mathbf{w}_\tau}$ the intersection of the half hyperplane $\mathbb{R}(E) \times \{c\mathbf{w}_\tau \mid c > 0\}$ with $\mathcal{S}_{\tau, \sigma}^{m-l'}$, namely

$$\mathcal{H}_{\mathbf{w}_\tau} = \left\{ \mathbf{w}_\sigma + c\mathbf{w}_\tau \in \mathcal{S}_{\tau, \sigma}^{m-l'} \mid c > 0 \text{ and } \mathbf{w}_\sigma \in \mathbb{R}(E) \times \{\mathbf{0}\} \subset \mathbb{R}(E) \times \mathcal{O}(F) \right\},$$

and let

$$\Omega_n^k(\mathbf{w}_\tau) = \left\{ \hat{\xi}_n \in \tau \text{ and } d\left(\frac{\hat{\xi}_n - \mathbf{x}^*}{\|\hat{\xi}_n - \mathbf{x}^*\|}, \mathcal{H}_{\mathbf{w}_\tau}\right) \leq \frac{1}{k} \right\}.$$

Proposition 9. *Let the stratum $\sigma = \mathcal{O}(E)$ of co-dimension $l (\geq 1)$ bound, in \mathbf{X}^m , the stratum $\tau = \mathcal{O}(E \cup F)$ of co-dimension $l' (< l)$. Assume that the Fréchet mean \mathbf{x}^* of μ lies in σ and that $\mathbf{w}_\tau \in \mathcal{S}_{\tau \setminus \sigma}^{l-l'}$ is given such that the inequality (17) corresponding to \mathbf{w}_τ is strict. Then,*

$$\lim_{k \rightarrow \infty} \mu \left(\limsup_n \Omega_n^k(\mathbf{w}_\tau) \right) = \lim_{k \rightarrow \infty} \mu \left(\bigcap_{l \geq 1} \bigcup_{n \geq l} \Omega_n^k(\mathbf{w}_\tau) \right) = 0.$$

Proof. For \mathbf{w}_τ as given in the proposition, let

$$\Omega_{\mathbf{w}_\tau} = \bigcap_{k \geq 1} \bigcap_{l \geq 1} \bigcup_{n \geq l} \Omega_n^k(\mathbf{w}_\tau).$$

Then, the set $\Omega_{\mathbf{w}_\tau}$ consists of points with the property that, for arbitrary $\epsilon > 0$, there exist arbitrarily large n such that $\hat{\xi}_n$ lies in τ and $(\hat{\xi}_n - \mathbf{x}^*)/\|\hat{\xi}_n - \mathbf{x}^*\|$ is within a distance ϵ of $\mathcal{H}_{\mathbf{w}_\tau}$. Since $\Omega_n^k(\mathbf{w}_\tau) \supseteq \Omega_n^{k+1}(\mathbf{w}_\tau)$, the required result is equivalent to showing that $\mu(\Omega_{\mathbf{w}_\tau}) = 0$.

Without loss of generality, we may assume that, restricted to $\Omega_{\mathbf{w}_\tau}$, $\hat{\xi}_n$ lie in τ for all n and $\mathbf{w}_n = (\hat{\xi}_n - \mathbf{x}^*)/\|\hat{\xi}_n - \mathbf{x}^*\| \rightarrow \mathbf{w}$ as $n \rightarrow \infty$ for some (random) unit vector $\mathbf{w} \in \mathcal{H}_{\mathbf{w}_\tau}$.

For the given $\mathbf{w}_\tau \in \mathcal{S}_{\tau \setminus \sigma}^{l-l'}$, each $\Psi_\tau(\xi_i, \mathbf{w}_\tau; \mathbf{x}^*)$ is a Euclidean random variable on $\mathbb{R}(E \cup F)$. Then, let

$$\hat{\xi}_n^{\mathbf{w}_\tau} = \frac{1}{n} \sum_{i=1}^n \Psi_\tau(\xi_i, \mathbf{w}_\tau; \mathbf{x}^*)$$

and write Ω_0 for the set consisting of points such that $\hat{\xi}_n^{\mathbf{w}_\tau}$ converges to

$$\int_{\mathbf{X}^m} \Psi_\tau(\mathbf{x}, \mathbf{w}_\tau; \mathbf{x}^*) d\mu(\mathbf{x}).$$

Then, it follows from the classical Law of Large Numbers that $\mu(\Omega_0) = 1$. Hence, restricted to $\Omega_{\mathbf{w}_\tau} \cap \Omega_0$, the assumption on the strictness of (17) implies that,

for some constant $c < 0$, there is an n_0 such that, for $n > n_0$, $\langle \mathbf{w}_\tau, \hat{\xi}_n^{\mathbf{w}_\tau} \rangle < c$. However, the assumption that $\mathbf{w}_n \rightarrow \mathbf{w} \in \mathcal{H}_{\mathbf{w}_\tau}$ implies that $\langle \mathbf{w}_\tau, \hat{\xi}_n - \mathbf{x}^* \rangle > 0$ for all sufficiently large n . Putting these two conclusions together, we have that, restricted to $\Omega_{\mathbf{w}_\tau} \cap \Omega_0$,

$$\langle \mathbf{w}_\tau, \hat{\xi}_n - \hat{\xi}_n^{\mathbf{w}_\tau} \rangle = \langle \mathbf{w}_\tau, \hat{\xi}_n - \mathbf{x}^* \rangle - \langle \mathbf{w}_\tau, \hat{\xi}_n^{\mathbf{w}_\tau} \rangle > -c > 0 \quad (25)$$

as $\langle \mathbf{w}_\tau, \mathbf{x}^* \rangle = 0$.

On the other hand, restricted to $\Omega_{\mathbf{w}_\tau} \cap \Omega_0$, $\hat{\xi}_n$ is in τ by the assumption made earlier. Then, it follows from (18) that

$$\hat{\xi}_n = \frac{1}{n} \sum_{i=1}^n \Phi_\tau(\xi_i; \hat{\xi}_n).$$

Thus, we can express the difference $\hat{\xi}_n - \hat{\xi}_n^{\mathbf{w}_\tau}$ as

$$\hat{\xi}_n - \hat{\xi}_n^{\mathbf{w}_\tau} = \frac{1}{n} \sum_{i=1}^n \left\{ \Phi_\tau(\xi_i; \hat{\xi}_n) - \Psi_\tau(\xi_i, \mathbf{w}_\tau; \mathbf{x}^*) \right\}. \quad (26)$$

Decompose $\mathbf{w}_n = (\mathbf{w}_n)_\sigma + (\mathbf{w}_n)^\perp$, where $(\mathbf{w}_n)_\sigma = P_\sigma(\mathbf{w}_n)$ and $(\mathbf{w}_n)^\perp = P_{\tau \setminus \sigma}(\mathbf{w}_n)$. Then, by Corollary 1(ii), the i th summand

$$\begin{aligned} & \Phi_\tau(\xi_i; \hat{\xi}_n) - \Psi_\tau(\xi_i, \mathbf{w}_\tau; \mathbf{x}^*) \\ &= \Phi_\tau(\xi_i; \hat{\xi}_n) - \Psi_\tau(\xi_i, \mathbf{w}_n; \mathbf{x}^*) + \Psi_\tau(\xi_i, \mathbf{w}_n; \mathbf{x}^*) - \Psi_\tau(\xi_i, \mathbf{w}_\tau; \mathbf{x}^*) \\ &= \Phi_\tau(\xi_i; \hat{\xi}_n) - \Psi_\tau(\xi_i, \mathbf{w}_n; \mathbf{x}^*) + \Psi_\tau(\xi_i, (\mathbf{w}_n)_\tau; \mathbf{x}^*) - \Psi_\tau(\xi_i, \mathbf{w}_\tau; \mathbf{x}^*), \end{aligned} \quad (27)$$

where $(\mathbf{w}_n)_\tau = (\mathbf{w}_n)^\perp / \|(\mathbf{w}_n)^\perp\| \in \mathcal{S}_{\tau \setminus \sigma}^{l-l'}$. As in the proof of Theorem 2, we may assume that the $\hat{\xi}_n$ are sufficiently close to \mathbf{x}^* that the carriers of the geodesics from $\hat{\xi}_n$ to ξ_i remain constant. Using the notation there, namely $\mathcal{A} = (A_0, \dots, A_k)$ and $\mathcal{B} = (B_0, \dots, B_k)$ for the common support $(\mathcal{A}, \mathcal{B})$ of the geodesics and W_s for $(A_s \cap E)_{\mathbf{x}^*}$ if $A_s \cap E \neq \emptyset$ and otherwise $(A_s \cap F)_{\mathbf{w}_n}$, then the s th set of components of $(j^{-1})(\Phi(\xi_i, \hat{\xi}_n))$ forms the vector $-\frac{\|B_s\|_{\xi_i}}{\|A_s\|_{\hat{\xi}_n}}(A_s)\hat{\xi}_n$ and, for $(j^{-1})(\Psi(\xi_i, \mathbf{w}_n; \mathbf{x}^*))$, the corresponding vector is $-\frac{\|B_s\|_{\xi_i}}{\|W_s\|}W_s$. When $A_s \cap E = \emptyset$, the projections of these two vectors onto $\mathbb{R}(F)$ are $-\|B_s\|_{\xi_i}$ times $\frac{(A_s \cap F)\hat{\xi}_n}{\|(A_s)\hat{\xi}_n\|}$ and $\frac{(A_s \cap F)\mathbf{w}_n}{\|(A_s \cap F)\mathbf{w}_n\|}$ respectively. Since $(A_s)\hat{\xi}_n = (A_s \cap F)\hat{\xi}_n = (A_s \cap F)\hat{\xi}_n - \mathbf{x}^*$, the latter two unit vectors in $\mathcal{S}_{\tau \setminus \sigma}^{l-l'}$ are identical. When $A_s \cap E \neq \emptyset$, these projections are both zero except, if $A_s \cap F \neq \emptyset$, for $\frac{(A_s \cap F)\hat{\xi}_n}{\|(A_s)\hat{\xi}_n\|}$ whose limit is then zero since $\|A_s\|_{\hat{\xi}_n} \neq 0$ but $(A_s \cap F)\hat{\xi}_n \rightarrow 0$. It follows that

$$\langle \mathbf{w}_\tau, \Phi_\tau(\xi_i; \hat{\xi}_n) - \Psi_\tau(\xi_i, \mathbf{w}_n; \mathbf{x}^*) \rangle \rightarrow 0, \quad \text{as } n \rightarrow \infty. \quad (28)$$

Moreover, since $\mathbf{w}^\perp = P_{\tau \setminus \sigma}(\mathbf{w}) \neq 0$, $\mathbf{w}_n \rightarrow \mathbf{w}$ implies that $(\mathbf{w}_n)_\tau \rightarrow \frac{\mathbf{w}^\perp}{\|\mathbf{w}^\perp\|} = \mathbf{w}_\tau$. Then, for sufficiently large n , it follows from a similar argument to that of the proof of Proposition 5 that

$$\begin{aligned} & \langle \mathbf{w}_\tau, \Psi_\tau(\xi_i, (\mathbf{w}_n)_\tau; \mathbf{x}^*) - \Psi_\tau(\xi_i, \mathbf{w}_\tau; \mathbf{x}^*) \rangle \\ & \approx \left\langle \mathbf{w}_\tau, (D\Psi_\tau(\xi_i, \mathbf{w}_\tau; \mathbf{x}^*)) \left(\arccos(\langle (\mathbf{w}_n)_\tau, \mathbf{w}_\tau \rangle) \frac{\mathbf{v}_n}{\|\mathbf{v}_n\|} \right) \right\rangle, \end{aligned}$$

where \mathbf{v}_n is the component of $(\mathbf{w}_n)_\tau - \mathbf{w}_\tau$ orthogonal to \mathbf{w}_τ , so that by Proposition 5

$$\langle \mathbf{w}_\tau, \Psi_\tau(\xi_i, (\mathbf{w}_n)_\tau; \mathbf{x}^*) - \Psi_\tau(\xi_i, \mathbf{w}_\tau; \mathbf{x}^*) \rangle \longrightarrow 0, \quad \text{as } n \rightarrow \infty. \quad (29)$$

Then, (26), (27), (28) and (29) together imply that, when it is restricted to $\Omega_{\mathbf{w}_\tau} \cap \Omega_0$, $\langle \mathbf{w}_\tau, \hat{\xi}_n - \hat{\xi}_n^{\mathbf{w}_\tau} \rangle \rightarrow 0$ as $n \rightarrow \infty$, contradicting (25). \square

When $l - l' = 1$, $\mathcal{S}_{\tau \setminus \sigma}^{l-l'}$ contains a single unit vector, so that we have the following special case. In particular, taking $l = 1$ and so $l' = 0$ recovers the result of Lemma 6 in [3] for the case of co-dimension one when \mathbf{X}^m is a tree space.

Corollary 5. *Let the stratum σ of co-dimension $l(\geq 1)$ bound, in \mathbf{X}^m , the stratum τ of co-dimension $l' = l - 1$. Assume that the Fréchet mean \mathbf{x}^* of μ lies in σ . If the inequality (17) corresponding to the unique $\mathbf{w}_\tau \in \mathcal{S}_{\tau \setminus \sigma}^{l-l'}$ is strict then, for all sufficiently large n , $\hat{\xi}_n$ cannot lie in τ .*

Thus, when $l - l' = 1$, the support of the limiting distribution of any appropriately scaled difference $\hat{\xi}_n - \mathbf{x}^*$ intersects the stratum $\mathbb{R}(E) \times \mathcal{O}(F)$ in the tangent cone to \mathbf{X}^m at \mathbf{x}^* only if the inequality (17) corresponding to the unique $\mathbf{w}_\tau \in \mathcal{S}_{\tau \setminus \sigma}^{l-l'}$ is an equality.

Similar to the case where $l - l' = 1$, Proposition 9 has the following consequence on the support of the limiting distribution when $l - l' > 1$, where $\mathcal{C}(\Theta)$ denotes the Euclidean cone on Θ .

Corollary 6. *Let the stratum $\sigma = \mathcal{O}(E)$ of co-dimension $l(\geq 2)$ bound, in \mathbf{X}^m , the stratum $\tau = \mathcal{O}(E \cup F)$ of co-dimension $l' \leq l - 2$. Assume that $\mathbf{x}^* \in \sigma$ is the Fréchet mean of μ . Then the support of the limiting distribution of an appropriately scaled difference $\hat{\xi}_n - \mathbf{x}^*$, if it meets the stratum $\mathbb{R}(E) \times \mathcal{O}(F)$ in the tangent cone to \mathbf{X}^m at \mathbf{x}^* , must be contained in $\mathbb{R}(E) \times \mathcal{C}(\Theta_{\tau, \sigma}(\mathbf{x}^*; \mu))$, where $\Theta_{\tau, \sigma}(\mathbf{x}^*; \mu)$ is defined by (20).*

Hence, the support of the limiting distribution of an appropriately scaled difference $\hat{\xi}_n - \mathbf{x}^*$ is contained in \mathcal{K}_μ where, for the closed sets

$$\mathcal{K}_{\mu, \tau} = \mathbb{R}(E) \times \mathcal{C}(\Theta_{\tau, \sigma}(\mathbf{x}^*; \mu)) \quad (30)$$

in the tangent cone to \mathbf{X}^m at \mathbf{x}^* ,

$$\mathcal{K}_\mu = \bigcup_{\tau \text{ co-bounds } \sigma} \mathcal{K}_{\mu, \tau} \quad (31)$$

and where we regard σ as co-bounding itself. Nevertheless, the following example shows that

- (i) if it is non-empty, $\mathbb{R}(E) \times \mathcal{C}(\Theta_{\tau, \sigma}(\mathbf{x}^*; \mu))$ is not necessarily an entire stratum $\mathbb{R}(E) \times \mathcal{O}(F)$;
- (ii) even if it is the entire stratum, the support of the limiting distribution of $\sqrt{n}(\hat{\xi}_n - \mathbf{x}^*)$ does not necessarily intersect that stratum; and

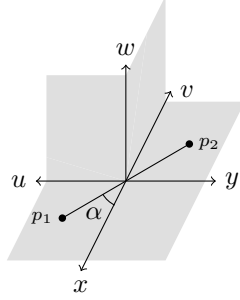


Figure 2: Probability measure μ on \mathbf{X}^2 has mass $1/2$ at p_1 and p_2 .

(iii) it is possible that the support of the limiting distribution, when restricted to the stratum, is only a subset of $\mathbb{R}(E) \times \mathcal{C}(\Theta_{\tau,\sigma}(\mathbf{x}^*; \mu))$.

Example. Consider the orthant space \mathbf{X}^2 comprising five orthants in \mathbb{R}^5 , illustrated in Figure 2, that was denoted Q_5 in [2]. Let μ have mass $1/2$ at the two points p_1 and p_2 equidistant from the cone point o along a geodesic through that point. Then its Fréchet mean is at the cone point and the sample Fréchet means always lie on this geodesic segment. On the other hand, writing $\tau_{x,y}$ for the top-dimensional stratum whose 1-dimensional bounding strata are τ_x and τ_y , indicated as x and y in Figure 2, it can be checked that

- (a) for any direction $\mathbf{w}_{\tau_{x,y}} \in \mathcal{S}_{\tau_{x,y} \setminus \{o\}}^2$, $\int_{\mathbf{X}^2} \Psi_{\tau_{x,y}}(p, \mathbf{w}_{\tau_{x,y}}; o) d\mu(p) = 0$, so that $\Theta_{\tau_{x,y}, \{o\}}(o; \mu) = \mathcal{S}_{\tau_{x,y} \setminus \{o\}}^2$, illustrating (ii) above;
- (b) for any direction $\mathbf{w}_{\tau_{x,u}} \in \mathcal{S}_{\tau_{x,u} \setminus \{o\}}^2$ with $\tan \mathbf{w}_{\tau_{x,u}} = u/x \leq \tan \alpha$,

$$\int_{\mathbf{X}^2} \Psi_{\tau_{x,u}}(p, \mathbf{w}_{\tau_{x,u}}; o) d\mu(p) = 0,$$

so that $\Theta_{\tau_{x,u}, \{o\}}(o; \mu) \supseteq \{\theta \in \mathcal{S}_{\tau_{x,u} \setminus \{o\}}^2 \mid \tan \theta \leq \alpha\}$;

- (c) for any direction $\mathbf{w}_{\tau_{x,u}} \in \mathcal{S}_{\tau_{x,u} \setminus \{o\}}^2$ with $\tan \mathbf{w}_{\tau_{x,u}} = u/x > \tan \alpha$,

$$\langle \mathbf{w}_{\tau_{x,u}}, \int_{\mathbf{X}^2} \Psi_{\tau_{x,u}}(p, \mathbf{w}_{\tau_{x,u}}; o) d\mu(p) \rangle < 0,$$

so that, combining with the conclusion of (b), $\Theta_{\tau_{x,u}, \{o\}}(o; \mu) = \{\theta \in \mathcal{S}_{\tau_{x,u} \setminus \{o\}}^2 \mid \tan \theta \leq \alpha\}$, illustrating (i) and (ii) above.

To describe the limiting distribution of $\sqrt{n}(\hat{\xi}_n - \mathbf{x}^*)$, where the Fréchet mean \mathbf{x}^* of μ lies in a stratum $\sigma = \mathcal{O}(E)$ of co-dimension $l \geq 0$, we continue to regard σ as co-bounding itself so that, in this case, the set F of additional axes in the ‘co-bounding’ stratum is empty. Moreover, we shall relate the form of the limiting distribution in the set (30) for each τ co-bounding σ to a limiting distribution of the Euclidean means of various Euclidean random variables depending on τ :

- (i) for $\tau = \sigma$, corresponding to the set $\mathbb{R}(E) \times \{\mathbf{0}\}$ in (30), the relevant Euclidean random variable is $\Phi_\sigma(\xi_1; \mathbf{x}^*)$;

- (ii) for $\tau \neq \sigma$ the relevant Euclidean random variable is $\Psi_\tau(\xi_1, \mathbf{w}_\tau; \mathbf{x}^*)$ where, if $l - l' > 1$, \mathbf{w}_τ is any chosen vector in $\text{int}(\Theta_{\tau,\sigma}(\mathbf{x}^*; \mu))$ if this set is not empty and, if $l - l' = 1$ with $\Theta_{\tau,\sigma}(\mathbf{x}^*; \mu) \neq \emptyset$, \mathbf{w}_τ is its unique element;
- (iii) we take the zero random variable otherwise.

Note that, by Proposition 8, different choices of \mathbf{w}_τ in the case $l - l' > 1$ of (ii) give random variables that are a.s. equal. Note also that, by Corollary 6, the random variables in the case (iii) play no role in the description of the limiting distribution so that they can be replaced by any other random variables. For simplicity, we denote the relevant random variable above in each case by $\tilde{\Psi}_\tau(\xi_1; \mathbf{x}^*)$. With this context and notation, for each τ let Z_τ be a random variable in $\mathbb{R}(E \cup F)$ with normal distribution $N(0, A^\top V_\tau A)$, where $V_\tau = \text{cov}(\tilde{\Psi}_\tau(\xi_1; \mathbf{x}^*))$,

$$A^{-1} = E \left[J^\top \text{diag}\{I_{m-l} - M_{\mathbf{x}^*}^\sigma(\xi_1), I_{l-l'}\} J \right].$$

Note that, since $M_{\mathbf{x}^*}^\sigma(\mathbf{x})$ is negative semi-definite, the above inverse is well defined when $E[M_{\mathbf{x}^*}^\sigma(\xi_1)]$ exists.

Theorem 4. *Let $\sigma = \mathcal{O}(E)$ be a stratum in \mathbf{X}^m of co-dimension $l(\geq 0)$. Assume that*

- (i) *the Fréchet mean \mathbf{x}^* of μ lies in σ ;*
- (ii) *$\mu(\mathcal{D}_{\mathbf{x}^*}) = 0$, where $\mathcal{D}_{\mathbf{x}^*}$ is given by Definition 8;*
- (iii) *$E[M_{\mathbf{x}^*}^\sigma(\xi_1)]$ exists, where $M_{\mathbf{x}^*}^\sigma(\mathbf{x})$ is given by (8);*
- (iv) *for any stratum τ in \mathbf{X}^m which co-bounds σ and has co-dimension $l' \leq l - 2$, if $\Theta_{\tau,\sigma}(\mathbf{x}^*; \mu) \neq \emptyset$ then $\text{int}(\Theta_{\tau,\sigma}(\mathbf{x}^*; \mu)) \neq \emptyset$.*

Then, if

$$\sqrt{n}\{\hat{\xi}_n - \mathbf{x}^*\} \xrightarrow{d} \eta, \quad \text{as } n \rightarrow \infty,$$

the random variable η on the tangent cone at \mathbf{x}^* has the following property: for any stratum $\tau = \mathcal{O}(E \cup F)$ of co-dimension $l'(\leq l)$ co-bounding σ , if $\mathbf{P}(\eta \in \mathbb{R}(E) \times \mathcal{O}(F)) > 0$ then, for Z_τ defined as above and \mathcal{K}_μ by (31),

$$\mathbf{P}(\eta \in B) = \mathbf{P}(Z_\tau \in B)$$

for any Borel set B contained in

$$\begin{cases} (\text{int}(\mathbb{R}(E) \times \mathcal{C}(\Theta_{\tau,\sigma}(\mathbf{x}^*; \mu)))) \setminus \partial\mathcal{K}_\mu & \text{if } l' \leq l - 2 \\ (\mathbb{R}(E) \times \mathcal{O}(F)) \setminus \partial\mathcal{K}_\mu & \text{if } l' = l - 1 \text{ or } l. \end{cases}$$

Proof. We assume that $l' \leq l - 2$. The case for $l' = l - 1$ can be similarly derived by noting Corollary 5, whereas for $l' = l$ the result can be derived directly by simplifying the following arguments.

Write $\Xi_\tau = \mathcal{O}(E) \times \mathcal{C}(\text{int}(\Theta_{\tau,\sigma}(\mathbf{x}^*; \mu)))$. Then, given $\hat{\xi}_n \in \tau$, $\hat{\xi}_n \in \Xi_\tau$ for sufficiently large n and, by Theorem 3, we also have

$$\begin{aligned} \sqrt{n}(\hat{\xi}_n - \mathbf{x}^*) &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ \Phi_\tau(\xi_i; \hat{\xi}_n) - \tilde{\Psi}_\tau(\xi_i; \mathbf{x}^*) \right\} \\ &\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ \tilde{\Psi}_\tau(\xi_i; \mathbf{x}^*) - \mathbf{x}^* \right\}. \end{aligned}$$

For any $\mathbf{x}' \in \tau$ and $\mathbf{x} \in \mathbf{X}^m$, denote the projection $P_{\tau \setminus \sigma}(\Phi_\tau(\mathbf{x}; \mathbf{x}'))$ of $\Phi_\tau(\mathbf{x}; \mathbf{x}')$ by $\Phi_{\tau \setminus \sigma}(\mathbf{x}; \mathbf{x}')$. Define $\tilde{\Psi}_{\tau \setminus \sigma}(\mathbf{x}; \mathbf{x}^*)$ similarly. Then, $\tilde{\Psi}_{\tau \setminus \sigma}(\mathbf{x}; \mathbf{x}^*) = \tilde{\Psi}_\tau(\mathbf{x}; \mathbf{x}^*) - \Phi_\sigma(\mathbf{x}; \mathbf{x}^*)$ by Corollary 2(ii). Since $\hat{\xi}_n$ is in Ξ_τ and converges to \mathbf{x}^* a.s., the result of Proposition 8 and the argument for the proof of Theorem 2 imply that, for any given \mathbf{x} and all sufficiently large n , $\Phi_{\tau \setminus \sigma}(\mathbf{x}; \hat{\xi}_n) = \tilde{\Psi}_{\tau \setminus \sigma}(\mathbf{x}; \mathbf{x}^*)$ a.s.. Hence, in particular, for sufficiently large n ,

$$P_{\tau \setminus \sigma}(\hat{\xi}_n - \mathbf{x}^*) = \frac{1}{n} \sum_{i=1}^n \Phi_{\tau \setminus \sigma}(\xi_i; \hat{\xi}_n)$$

is a.s. the Euclidean mean of $\tilde{\Psi}_{\tau \setminus \sigma}(\xi_1; \mathbf{x}^*), \dots, \tilde{\Psi}_{\tau \setminus \sigma}(\xi_n; \mathbf{x}^*)$. It follows that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ \Phi_{\tau \setminus \sigma}(\xi_i; \hat{\xi}_n) - \tilde{\Psi}_{\tau \setminus \sigma}(\xi_i; \mathbf{x}^*) \right\} 1_{\Xi_\tau}(\hat{\xi}_n) \xrightarrow{P} 0.$$

Thus, the limiting distribution of $\sqrt{n}(\hat{\xi}_n - \mathbf{x}^*) 1_{\Xi_\tau}(\hat{\xi}_n)$ is the same as that of

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ \Phi_\sigma(\xi_i; \hat{\xi}_n) - \tilde{\Psi}_\sigma(\xi_i; \mathbf{x}^*) \right\} 1_{\Xi_\tau}(\hat{\xi}_n) + \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ \tilde{\Psi}_\tau(\xi_i; \mathbf{x}^*) - \mathbf{x}^* \right\} 1_{\Xi_\tau}(\hat{\xi}_n).$$

Since $\tilde{\Psi}_\sigma(\xi_i; \mathbf{x}^*) = \Phi_\sigma(\xi_i; \mathbf{x}^*)$, Proposition 4 implies that the limiting distribution of $\sqrt{n}(\hat{\xi}_n - \mathbf{x}^*) 1_{\Xi_\tau}(\hat{\xi}_n)$ is equal to that of

$$\sqrt{n} P_\sigma(\hat{\xi}_n - \mathbf{x}^*) 1_{\Xi_\tau}(\hat{\xi}_n) \frac{1}{n} \sum_{i=1}^n M_{\mathbf{x}^*}^\sigma(\xi_i) + 1_{\Xi_\tau}(\hat{\xi}_n) \frac{1}{\sqrt{n}} \sum_{i=1}^n \left\{ \tilde{\Psi}_\tau(\xi_i; \mathbf{x}^*) - \mathbf{x}^* \right\}.$$

Hence, by (23), the required result follows from a similar argument to that used in [2] and [3]. \square

As for $\Theta_\sigma(\mathbf{x}^*; \mu)$ defined by (21), the convexity in \mathbf{w} of the directional derivative $D(d_{\mathbf{x}}^2)(\mathbf{w})$ implies that \mathcal{K}_μ is a convex subset of the tangent cone to \mathbf{X}^m at \mathbf{x}^* . This, together with the structure of an orthant space, implies that the result of Theorem 4 refers to the behaviour of the limiting distribution only within the interior of \mathcal{K}_μ . Its behaviour at the boundaries will depend on how these sets relate to each other and on the shape of the boundary $\partial \mathcal{K}_\mu$.

The assumption in Theorem 4 that $\mu(\mathcal{D}_{\mathbf{x}^*}) = 0$ ensures that we are able to employ the so-called delta method for the approximate probability distribution of a function of an asymptotically normal statistical estimator. In principle, it is possible to relax this assumption by using directional derivatives and combining that with the use of the law of the total probability. However, it is clear from the discussion leading to the definition of $\mathcal{D}_{\mathbf{x}^*}$ that its structure, although conceptually straightforward, is generally more complex than will admit a simple algebraic representation, and the ensuing results will consequently depend heavily on the behaviour of μ on $\mathcal{D}_{\mathbf{x}^*}$.

To observe special cases of Theorem 4, let $\sigma = \mathcal{O}(E)$ be a stratum in \mathbf{X}^m of co-dimension $l(\geq 0)$ in which the Fréchet mean \mathbf{x}^* of μ lies, assume that the conditions of Theorem 4 are satisfied and write

$$l(\mu) = \inf\{l' \mid l' = \text{co-dimension of } \tau, \text{ where } \Theta_{\tau, \sigma}(\mathbf{x}^*; \mu) \neq \emptyset\},$$

where we assume that $l(\mu) = l$ if there is no τ with co-dimension $l' < l$ which satisfies the above required condition. We assume further that, for $\tau = \mathcal{O}(E \cup F)$ of co-dimension $l(\mu)$ co-bounding σ and, if $l(\mu) < l$, with $\Theta_{\tau,\sigma}(\mathbf{x}^*; \mu) \neq \emptyset$, $V_\tau = \text{cov}(\tilde{\Psi}_\tau(\boldsymbol{\xi}_1; \mathbf{x}^*))$ is of full rank $m - l(\mu)$. Then, it is clear from the proof of Theorem 4 that $\mathbf{P}(\eta \in \mathbb{R}(E) \times \mathcal{O}(F)) > 0$.

Case $l(\mu) = l$: in this case, $\mathcal{K}_\mu = \mathbb{R}(E)$ and the support of the distribution of η is contained in the tangent space of σ . Then, Theorem 4 says that η is a normal random variable with mean zero and covariance matrix $A_\sigma^\top \text{cov}(\Phi_\sigma(\boldsymbol{\xi}_1; \mathbf{x}^*)) A_\sigma$, where

$$A_\sigma = \{I_{m-l} - E[M_{\mathbf{x}^*}^\sigma(\boldsymbol{\xi}_1)]\}^{-1}.$$

This generalises the limiting distribution of $\sqrt{n}\{\hat{\boldsymbol{\xi}}_n - \mathbf{x}^*\}$ when \mathbf{x}^* lies in a top-dimensional stratum of a tree space obtained in [3].

Case $l(\mu) = l - 1$ so that $l \geq 1$: if $\tau = \mathcal{O}(E \cup F)$ is a stratum of co-dimension $l' = l - 1$ such that $\Theta_{\tau,\sigma}(\mathbf{x}^*; \mu) \neq \emptyset$, then F contains only one axis. By taking the Borel set $B = \mathbb{R}(E) \times \mathcal{O}(F)$, we see that $\mathbf{P}(\eta \in \mathbb{R}(E) \times \mathcal{O}(F)) = 1/2$ since the corresponding Z_τ is a normal random variable in \mathbb{R}^{m-l+1} with mean zero. Hence, there are at most two strata of co-dimension $l(\mu)$ co-bounding σ on which infinitely many $\hat{\boldsymbol{\xi}}_n$ lie. Moreover, in the case of there being only one such a stratum, $\mathbf{P}(\eta \in \sigma) = 1/2$ and, in case of two such strata, $\mathbf{P}(\eta \in \sigma) = 0$.

Case that $0 \leq l(\mu) < l$, that there is a single $\tau_0 = \mathcal{O}(E \cup F_0)$ such that the co-dimension of τ_0 is $l(\mu)$ and that $\Theta_{\tau_0,\sigma}(\mathbf{x}^; \mu) = \mathcal{S}_{\tau_0 \setminus \sigma}^{l-l(\mu)}$:* in this case, we have the following full description of the distribution of η in terms of ϕ_{τ_0} , the probability density function of the random variable Z_{τ_0} defined prior to Theorem 4. We first note that, since \mathcal{K}_μ defined by (31) is convex and closed, the result of Proposition 8 implies that, in this case,

$$\mathcal{K}_\mu = \bigcup_{F \subseteq F_0} \mathbb{R}(E) \times \mathcal{O}(F).$$

Then, we extend the projection map P to $\mathbb{R}(E \cup F_0)$ in an obvious fashion and, for any $\tau = \mathcal{O}(E \cup F)$, where $F \subseteq F_0$, and any $\mathbf{z} \in \mathbb{R}(E \cup F_0)$, write $\mathbf{z}_\tau = P_\tau(\mathbf{z})$ and $\mathbf{z}_{\tau_0 \setminus \tau} = P_{\tau_0 \setminus \tau}(\mathbf{z}) = \mathbf{z} - \mathbf{z}_\tau$.

Proposition 10. *Under the above assumptions and notation, the limiting distribution of $\sqrt{n}\{\hat{\boldsymbol{\xi}}_n - \mathbf{x}^*\}$ is given as follows: for any $\tau = \mathcal{O}(E \cup F)$, where $F \subseteq F_0$, and any Borel subset $B \subseteq \mathbb{R}(E) \times \mathcal{O}(F)$,*

$$\mathbf{P}(\eta \in B) = \int_B \psi_F(\mathbf{z}_\tau) d\mathbf{z}_\tau,$$

where

$$\psi_F(\mathbf{z}_\tau) = \int_{-\infty}^0 \phi_{\tau_0}(\mathbf{z}) d\mathbf{z}_{\tau_0 \setminus \tau}.$$

The special case that $l(\mu) = l - 1$ of this Proposition, together with the comments in the previous two paragraphs, generalises the limiting distribution of $\sqrt{n}\{\hat{\boldsymbol{\xi}}_n - \mathbf{x}^*\}$ when \mathbf{X}^m is a tree space and \mathbf{x}^* lies in a stratum of co-dimension one obtained in [3].

Proof. By Theorem 4, we only need to consider the case where $F \neq F_0$.

Assume that $\tau = \mathcal{O}(E \cup F)$ has co-dimension l' and fix $\mathbf{w}_\tau \in S_{\tau \setminus \sigma}^{l-l'}$. Since the number of hyper-surfaces determining $\mathcal{D}_{\mathbf{x}} \cap \tau_0$ is finite for any $\mathbf{x} \in \mathbf{X}^m$, the geodesics from $\mathbf{x}^*(\lambda, \mathbf{w}_{\tau_0}) = \mathbf{x}^* + \lambda \mathbf{w}_{\tau_0}$ to \mathbf{x} have the same support for all $\mathbf{w}_{\tau_0} \in S_{\tau_0 \setminus \sigma}^{l-l(\mu)}$ sufficiently close to \mathbf{w}_τ and all sufficiently small $\lambda > 0$. For such \mathbf{w}_{τ_0} and λ , by Definition 11, Corollary 2(i), Propositions 7 and 8, the sequence $\mathcal{A} = (A_0, \dots, A_k)$ in the support $(\mathcal{A}, \mathcal{B})$ of the geodesics from $\mathbf{x}^*(\lambda, \mathbf{w}_{\tau_0}) = \mathbf{x}^* + \lambda \mathbf{w}_{\tau_0}$ to ξ_1 a.s. has the property that, if $i > 0$ and if $(A_i \cap E) = \emptyset$, then A_i consists of a single axis, so that the axis $(A_i)_{\mathbf{x}^*(\lambda, \mathbf{w}_{\tau_0})} / \|A_i\|_{\mathbf{x}^*(\lambda, \mathbf{w}_{\tau_0})} \in F_0$ is independent of the value of λ . This, together with the fact implied by Corollary 2(ii) that, if $(A_i \cap E) \neq \emptyset$, $(A_i)_{\mathbf{x}^*(\lambda, \mathbf{w}_{\tau_0})} \rightarrow (A_i)_{\mathbf{x}^*}$ as $\lambda \rightarrow 0$, shows that, with probability one, each $(A_i)_{\mathbf{x}^*(\lambda, \mathbf{w}_{\tau_0})} / \|A_i\|_{\mathbf{x}^*(\lambda, \mathbf{w}_{\tau_0})}$ in the expression (3) for $\Phi_{\tau_0}(\xi_1; \mathbf{x}^* + \lambda \mathbf{w}_{\tau_0})$ is a continuous function at \mathbf{x}^* in the corresponding Euclidean space. It follows that

$$\lim_{\substack{\mathbf{w}_{\tau_0} \rightarrow \mathbf{w}_\tau \\ \lambda \rightarrow 0+}} \Phi_{\tau_0}(\xi_1; \mathbf{x}^* + \lambda \mathbf{w}_{\tau_0})$$

exists a.s. and so, in particular,

$$\lim_{\mathbf{w}_{\tau_0} \rightarrow \mathbf{w}_\tau} \lim_{\lambda \rightarrow 0+} \Phi_{\tau_0}(\xi_1; \mathbf{x}^* + \lambda \mathbf{w}_{\tau_0}) = \lim_{\lambda \rightarrow 0+} \lim_{\mathbf{w}_{\tau_0} \rightarrow \mathbf{w}_\tau} \Phi_{\tau_0}(\xi_1; \mathbf{x}^* + \lambda \mathbf{w}_{\tau_0}) \quad a.s.$$

Thus, the definition of Ψ gives

$$\lim_{\mathbf{w}_{\tau_0} \rightarrow \mathbf{w}_\tau} \Psi_{\tau_0}(\xi_1, \mathbf{w}_{\tau_0}; \mathbf{x}^*) = \lim_{\lambda \rightarrow 0+} \lim_{\mathbf{w}_{\tau_0} \rightarrow \mathbf{w}_\tau} \Phi_{\tau_0}(\xi_1; \mathbf{x}^* + \lambda \mathbf{w}_{\tau_0}) \quad a.s.$$

Since the limit on the right hand side exists, to find it, we take a particular path for \mathbf{w}_{τ_0} to approach \mathbf{w}_τ : $\mathbf{w}_{\tau_0} = \sin \alpha \mathbf{w}^\perp + \cos \alpha \mathbf{w}_\tau$, where $\langle \mathbf{w}^\perp, \mathbf{w}_\tau \rangle = 0$ and $\|\mathbf{w}^\perp\| = 1$. Then, writing $\beta = \lambda \sin \alpha$, we have

$$\begin{aligned} & \lim_{\mathbf{w}_{\tau_0} \rightarrow \mathbf{w}_\tau} \Phi_{\tau_0}(\xi_1, \mathbf{x}^* + \lambda \mathbf{w}_{\tau_0}) \\ &= \lim_{\alpha \rightarrow 0+} \Phi_{\tau_0}(\xi_1, \mathbf{x}^*(\lambda, \mathbf{w}_\tau) + \lambda(\sin \alpha \mathbf{w}^\perp + (\cos \alpha - 1) \mathbf{w}_\tau)) \\ &= \lim_{\beta \rightarrow 0+} \Phi_{\tau_0}(\xi_1, \mathbf{x}^*(\lambda, \mathbf{w}_\tau) + \beta \mathbf{w}^\perp) \\ &= \Psi_{\tau_0}(\xi_1, \mathbf{w}^\perp; \mathbf{x}^*(\lambda, \mathbf{w}_\tau)) \quad a.s., \end{aligned}$$

where the second equality follows from Corollary 1(ii). Hence, it follows from Corollary 2(ii) that

$$P_\tau \left(\lim_{\mathbf{w}_{\tau_0} \rightarrow \mathbf{w}_\tau} \Psi_{\tau_0}(\xi_1, \mathbf{w}_{\tau_0}; \mathbf{x}^*) \right) = \lim_{\lambda \rightarrow 0+} \Phi_\tau(\xi_1; \mathbf{x}^*(\lambda, \mathbf{w}_\tau)) = \Psi_\tau(\xi_1, \mathbf{w}_\tau; \mathbf{x}^*) \quad a.s.$$

as $\mathbf{x}^*(\lambda, \mathbf{w}_\tau) \in \tau$. That is,

$$P_\tau \left(\lim_{\mathbf{w}_{\tau_0} \rightarrow \mathbf{w}_\tau} \tilde{\Psi}_{\tau_0}(\xi_1; \mathbf{x}^*) \right) = \tilde{\Psi}_\tau(\xi_1; \mathbf{x}^*) \quad a.s. \quad (32)$$

Since $\hat{\xi}_n$ will lie in \mathcal{K}_μ for sufficiently large n a.s., without loss of generality, we assume that it is true for all n . Let $\hat{\xi}_n^\tau$ denote the sample Euclidean mean of $\tilde{\Psi}_\tau(\xi_1; \mathbf{x}^*), \dots, \tilde{\Psi}_\tau(\xi_1; \mathbf{x}^*)$. Then, by Corollary 4, $\hat{\xi}_n^\tau \rightarrow \mathbf{x}^* \in \mathbb{R}(E \cup F)$ a.s.

and application of (32) gives $\left(\hat{\xi}_n^\tau\right)_{\tau \setminus \sigma} = \left(\hat{\xi}_n^{\tau_0}\right)_{\tau \setminus \sigma}$. Now, the argument in the proof of Theorem 4 implies that, for all sufficiently large n ,

$$1_\tau(\hat{\xi}_n) \left(\hat{\xi}_n\right)_{\tau \setminus \sigma} = 1_\tau(\hat{\xi}_n) \left(\hat{\xi}_n - \mathbf{x}^*\right)_{\tau \setminus \sigma} = 1_\tau(\hat{\xi}_n) \left(\hat{\xi}_n^\tau\right)_{\tau \setminus \sigma}$$

so that, for all sufficiently large n ,

$$1_\tau(\hat{\xi}_n) \left(\hat{\xi}_n\right)_{\tau \setminus \sigma} = 1_\tau(\hat{\xi}_n) \left(\hat{\xi}_n^{\tau_0}\right)_{\tau \setminus \sigma}.$$

However, for all sufficiently large n , the fact that $\hat{\xi}_n \in \tau$ is equivalent to the fact that, for sufficiently large n , $\left(\hat{\xi}_n^{\tau_0}\right)_{\tau \setminus \sigma}$ lies in $\mathcal{O}(F)$ and $\left(\hat{\xi}_n^{\tau_0}\right)_{\tau_0 \setminus \tau}$ lies in the negative orthant of $\mathbb{R}(F_0 \setminus F)$, i.e. $-\left(\hat{\xi}_n^{\tau_0}\right)_{\tau_0 \setminus \tau} \in \mathcal{O}(F_0 \setminus F)$. Hence, we can re-express the above equality as

$$1_\tau(\hat{\xi}_n) \left(\hat{\xi}_n\right)_{\tau \setminus \sigma} = 1_{\mathcal{O}(F)} \left(\left(\hat{\xi}_n^{\tau_0}\right)_{\tau \setminus \sigma}\right) 1_{\mathcal{O}(F_0 \setminus F)} \left(-\left(\hat{\xi}_n^{\tau_0}\right)_{\tau_0 \setminus \tau}\right) \left(\hat{\xi}_n^{\tau_0}\right)_{\tau \setminus \sigma}.$$

The required result then follows by a slight modification to the proof of Theorem 4. \square

In fact, the argument for the proof of Proposition 10, in particular (32), also shows that, if $\tau = \mathcal{O}(E \cup F)$ has co-dimension greater than $l(\mu)$ and if $\mathbb{R} \times \Theta_{\tau, \sigma}(\mathbf{x}^*; \mu)$ is contained in the interior of \mathcal{K}_μ , then $\mathbf{P}(\eta \in \mathbb{R} \times \mathcal{O}(F)) = 0$.

Acknowledgements. The authors are indebted to Megan Owen for her continuing helpful discussions, following her collaboration in [2] and [3]. The second author acknowledges funding from the Engineering and Physical Sciences Research Council.

References

- [1] M. Bacak (2014). Computing medians and means in Hadamard spaces, *SIAM J. Optimiz.* **24**, 1542-1566.
- [2] D. Barden, H. Le and M. Owen (2013). Central limit theorems for Fréchet means in the space of phylogenetic trees, *Electron. J. Probab.* **18**, no. 25.
- [3] D. Barden, H. Le and M. Owen (2014). Limiting behaviour of Fréchet means in the space of phylogenetic trees, arXiv:1409.7602v1 [math.PR].
- [4] R. Bhattacharya and V. Patrangenaru (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds-II, *Ann. Statist.* **33**, 1225-1259.
- [5] L.J. Billera, S.P. Holmes and K. Vogtmann (2001). Geometry of the space of phylogenetic trees, *Advances in Applied Mathematics* **27**, 733-767.
- [6] M.R. Bridson and A. Haefliger (1999). *Metric Spaces of Non-positive Curvature*. Springer-Verlag, Berlin/New York.

- [7] B. Burago, Y. Burago and S. Ivanov (2001). *A Course in Metric Geometry*. American Mathematical Society.
- [8] M. Goresky and R. MacPherson (1980). *Stratified Morse Theory*. Springer-Verlag, Berlin/New York.
- [9] S. Holmes (2003). Statistics for phylogenetic trees, *Theoretical Population Biology* **63**, 17-32.
- [10] T. Hotz, S. Huckemann, H. Le, J.S. Marron, J.C. Mattingly, E. Miller, J. Nolen, M. Owen, V. Patrangenaru and S. Skwerer (2013). Sticky central limit theorems on open books, *Ann. Appl. Probab.* **23**, 2238-2258.
- [11] S. Huckemann, J.C. Mattingly, E. Miller and J. Nolen (2015). Sticky central limit theorems at isolated hyperbolic planar singularities, *Electron. J. Probab.* **20**, no. 78.
- [12] M. Kendall and C. Colijn (2015). A tree metric using structure and length to capture distinct phylogenetic signals, arXiv:1507.05211 [q-bio.PE].
- [13] W.S. Kendall and H. Le (2011). Limit theorems for empirical Fréchet means of independent and non-identically distributed manifold-valued random variables, *Brazilian Journal of Probability and Statistics* **25**, 323-352.
- [14] B. Kleiner (1999). The local structure of length spaces with curvature bounded above, *Math. Z.* **231**, 409-456.
- [15] E. Miller, M. Owen and S. Provan (2015). Polyhedral computational geometry for averaging metric phylogenetic trees, *Advances in Applied Mathematics* **68**, 51-91.
- [16] I. Nikolaev (1995). The tangent cone of an Aleksandrov space of curvature $\leq K$, *Manuscripta Math.* **86**, 137-147.
- [17] T.M.W. Nye (2011). Principal components analysis in the space of phylogenetic trees, *Ann. Statist.* **39**, 2716-2739.
- [18] T.M.W. Nye (2015). Convergence of random walks to Brownian motion in phylogenetic tree-space, arXiv:1508.02906 [q-bio.PE].
- [19] M. Owen (2011). Computing geodesic distances in tree space, *SIAM J. Discrete Math.* **25**, 1506-1529.
- [20] R. Thom (1969). Ensembles et morphismes stratifiés, *Bull. Amer. Math. Soc.* **75**, 240-284.
- [21] K. Vogtmann (2007). Geodesics in the space of trees, available at www.math.cornell.edu/~vogtmann/papers/TreeGeodesicss/index.html.
- [22] H. Ziezold (1977). On expected figures and a strong law of large numbers for random elements in quasi-metric spaces, *Trans. 7th Prague Conf. Info. Theory, Stat. Dec. Functions, Rand. Proc.* **A**, 591-602.